Design Principles of Mammalian Signaling Networks: Emergent Properties at Modular and Global Scales

by

Jason W. Locasale

B.A. Chemistry, B.A. Physics Rutgers University, Rutgers College 2003

M.S. Chemistry University of California, Berkeley 2005

SUBMITTED TO THE DEPARTMENT OF BIOLOGICAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
[Ph.D.]
DOCTER OF SCIENCE IN BIOLOGICAL ENGINEERING

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
[June 2008]
May 2008

Signature of Author: _____

Department of Biological Engineering May 2008

Certified by: _____

Arup K. Chakraborty, Ph.D. Professor of Biological Engineering, Robert T. Haslam Professor of Chemical Engineering, Professor of Chemistry: Thesis Supervisor

Certified by: _____

Alan J. Grodzinsky, Ph.D. Chair of Graduate Committee

This thesis has been accepted by a committee of the Biological Engineering Department as follows:

Professor Douglas A. Lauffenburger

Uncas and Helen Whitaker Professor and Director of Biological Engineering, Professor of Chemical Engineering, Professor of Biology

Thesis Committee Chairman

Professor Arup K. Chakraborty

Professor of Biological Engineering, Robert T. Haslam Professor of Chemical Engineering, Professor of Chemistry

Thesis Supervisor

Professor Forest M. White

Mitsui Career Development Associate Professor of Biological Engineering

Professor Mehran H. Kardar

Professor of Physics

Abstract

This thesis utilizes modeling approaches rooted in statistical physics and physical chemistry to investigate several aspects of cellular signal transduction at both the modular and global levels. Design principles of biological networks and cell signaling processes pertinent to disease progression emerge from these studies. It is my hope that knowledge of these principles may provide new mechanistic insights and conceptual frameworks for thinking about therapeutic intervention into diseases such as cancer and diabetes that arise from aberrant signaling. Areas of interest have emphasized the role of scaffold proteins in protein kinase cascades, modeling relevant biophysical processes related to T cell activation, design principles of signal transduction focusing on multisite phosphorylation, quantifying the notion of signal duration and the time scale dependence of signal detection, and entropy based models of network architecture inferred from proteomics data. These problems are detailed below.

The assembly of multiple signaling proteins into a complex by a scaffold protein guides many cellular decisions. Despite recent advances, the overarching principles that govern scaffold function are not well understood. We carried out a computational study using kinetic Monte Carlo simulations to understand how spatial localization of kinases on a scaffold may regulate signaling under different physiological condition. Our studies identify regulatory properties of scaffold proteins that allow them to both amplify and attenuate incoming signals in different biological contexts. In a further, supplementary study, simulations also indicate that a major effect that scaffolds exert on the dynamics of cell signaling is to control how the activation of protein kinases is distributed over time[2]. Scaffolds can influence the timing of kinase activation by allowing for kinases to become activated over a broad range of times, thus allowing for signaling across a broad spectrum of time scales.

T cells orchestrate the adaptive immune response and are central players in maintenance of functioning immune system. Recent studies have reported that T cells can integrate signals between interrupted encounters with Antigen Presenting Cells (APCs) in such a way that the process of signal integration exhibits a form of memory. We carried out a computational study using a simple mathematical model of T cell activation to investigate the ramifications of interrupted T cell-APC contacts on signal integration. We considered several mechanisms of how signal integration at these time scales may be achieved. In another study, we investigated the role of spatially localizing signaling components of the T cell signaling pathway into a structure known as the immunological synapse. We constructed a minimal mathematical model that offers a mechanism for how antigen quality can regulate signaling dynamics in the immunological synapse

These studies involving the analysis of signaling dynamics led us to investigate how differences in signal duration might be detected. Signal duration (e.g. the time scales over which an active signaling intermediate persists) is a key regulator of biological decisions in myriad contexts such as cell growth, proliferation, and developmental lineage commitments. Accompanying differences in signal duration are numerous downstream biological processes that require multiple steps of biochemical regulation. We present an analysis that investigates how simple biochemical motifs that involve multiple stages of regulation can be constructed to differentially process signals

that persist at different time scales[3]. Topological features of these networks that allow for different frequency dependent signal processing properties are identified.

One role of multisite phosphorylation in cell signaling is also investigated. The utilization of multiple phosphorylation sites in regulating a biological response is ubiquitous in cell signaling. If each site contributes an additional, equivalent binding site, then one consequence of an increase in the number of phosphorylations may be to increase the probability that, upon disassociation, a ligand immediately rebinds to its receptor. How such effects may influence cell signaling systems is not well understood. A self-consistent integral equation formalism for ligand rebinding, in conjunction with Monte Carlo simulations, was employed to further investigate the effects of multiple, equivalent binding sites on shaping biological responses.

Finally, this thesis also seeks to investigate cell signaling at a global scale. Advances in Mass Spectrometry based phosphoproteomics have allowed for the real-time quantitative monitoring of entire proteomes as signals propagate through complex networks in response to external signals. The trajectories of as many as 222 phosphorylated tyrosine sites can be simultaneously and reproducibly monitored at multiple time points. We develop and apply a method using the principle of maximum entropy to infer a model of network connectivity of these phosphorylation sites. The model predicts a core structure of signaling nodes, affinity dependent topological features of the network, and connectivity of signaling nodes that were hitherto unassociated with the canonical growth factor signaling network.

Our combined results illustrate many complexities in the broad array of control properties that emerge from the physical effects that constrain signal propagation on complex biological networks. It is the hope of this work that these studies bring coherence to seemingly paradoxical observations and suggest that cells have evolved design rules that enable biochemical motifs to regulate widely disparate cellular functions.

Table of Contents:

Chapter 5 – Allovalency Revisited: An Analysis of Multisite Phosphorylation and Substrate Rebinding.....128

Acknowledgements

It has been many years since I first decided to pursue doctoral studies in graduate school and, hopefully, a career in academics and scientific research; and, through these years, I have many to thank.

I thank my early intellectual and scientific mentors without whom I would not be in the position I am in today. I thank my high school science teachers and college professors for sparking a curiosity about the natural world and instilling in me the confidence to pursue independent research.

I thank my family, particularly my brother and my mom for their unwavering support despite dealing with my difficulties and at times, my inattention to them. I thank my friends, especially those in graduate school with whom I was able to share my own experiences, learn from, and grow both intellectually and emotionally – particularly to Roger, Brad, Medha, Marvin, Galia, Sarai, Christie, among many others.

I thank my thesis committee at UC-Berkeley: Professors Phillip Geissler, David Chandler, George Oster, and Jay Groves for their help in the beginning of graduate school.

I thank my scientific collaborators, particularly Dr. Andrey Shaw and his laboratory for helpful insights into biological problems. Much of the work in this thesis was inspired by experimental findings originating from his lab.

I thank my current Biological Engineering thesis committee: Mehran Kardar, Forest White, Doug Lauffenburger, and Arup for their support and inspiration. Each of them has a unique and complementary skill set that allowed me to learn much from each of them. I especially thank Doug for welcoming me into his department and being so supportive and accommodating.

I thank my thesis advisor, Arup Chakraborty for most of all establishing a research group both here at MIT and at Berkeley full of talented and creative graduate students and postdocs. I especially thank my early mentors in Arup's group for teaching me so much; particularly, Shuyan Qi, Jian Liu, Baron Peters, Jayajit Das, and Mike Hagan.

Finally, I thank several financial institutions for generous support. Support from the UC Berkeley Regents' fellowship, LBNL, UCB teaching assistantships, project grant (PO1 AI071195-01), and the NIH Director's Pioneer Award are each gratefully acknowledged.

Chapter 1 – Introduction

Chapter 1.1. Background

Signal transduction is the science of cellular communication – cells detect environmental signals in the form of antigens, hormones, stresses, growth factors, among many others, and integrate them into appropriate phenotypic responses such as cell growth, proliferation, division, or apoptosis (programmed cell death). Higher organisms have evolved complex biochemical networks endowed with many regulatory mechanisms to successfully detect, process, and convert these external, receptor tyrosine kinase (RTK) mediated, signals into different biological outcomes. Since aberrations in the regulation of signal transduction are the hallmarks of diseases such as cancer[4] or autoimmune disorders (e.g. diabetes)[5], there is a tremendous motivation to study the underlying principles that comprise these crucial processes.

Recent advances in several technologies, including but not limited to imaging, genetics, biochemistry, and high throughput techniques such as microarrays and mass spectrometry have uncovered a wealth of new data on many molecular and quantitative aspects of cell signaling[4, 6-9]. New mathematical and computational methodologies are undoubtedly required to consolidate this abundance of data into new theoretical frameworks to allow for better mechanistic understanding and new predictive capabilities[10].

Physical chemistry, with part of its foundations in the application of statistical physics and chemical kinetics to study complex molecular systems provides a natural framework to study the biochemical processes that are pertinent to cell signaling. However, the theoretical frameworks common to chemical physics are approaches that have not been widely utilized in the biological and medical communities. This is largely

due to a lack of communication between the two large, seemingly disparate, disciplines. In this thesis, I develop and apply methodologies rooted in statistical physics, information theory, and physical chemistry to address several problems in cell signaling in higher eukaryotic organisms. In carrying out such an undertaking, it is my hope to introduce physical scientists to new problems in molecular biology that are both fundamentally interesting and of broad societal interest due to their enormous medical applications. Also in perhaps a more ambitious undertaking, this thesis aims to illustrate how mathematical and computational frameworks rooted in physical science can be used to address problems that are of broad interest to the biology and medical communities. Ultimately, these modeling efforts lead to new testable biological hypotheses.

Much of this thesis concerns analysis of isolated biochemical modules or pathways that are separately analyzed form larger signaling systems– signaling at the modular scale. The emergent (i.e. systems level) properties that arise from the collective interactions among the many molecular components within these modules are characterized. However, the final chapter reports work on modeling and analyzing the behavior of a large signaling network comprising of N = 222 experimentally measured phosphorylation sites – signaling at the global scale across an entire signaling network. Analyses of the mechanisms of signal transduction, at both these scales, I believe, are complementary.

Design principles are detailed for how scaffold proteins, proteins that assemble multiple components of a signal transduction pathway into close proximity, confer intrinsic regulatory properties to protein kinase cascades by introducing physical constraints into the process of signal transduction(Chapters 2, 3). Also, the issue of how

cells biochemically regulate and detect the time scale dependence of incoming signals is discussed (Chapters 3, 4, 5). Another problem concerns functional aspects of multisite phosphorylation – a ubiquitous motif in cell signaling systems (Chapter 5). Signaling in the immune system, particularly aspects of T cell activation, the process by which a T cell recognizes the presence of a pathogen and, through cell signaling, initiates the onset of an adaptive immune response, is also considered (Chapters 6,7). Work involving the analysis of the behavior of global signaling networks as measured by phosphoproteomics (experiments that probe the dynamics of entire signaling networks across a whole cell) is discussed at the end (Chapter 8). Finally, conclusions and a future outlook are mentioned (Chapter 9).

Chapter 1.2. The scope of the thesis

Here, I provide a more detailed overview of the work contained in subsequent chapters of this thesis. The first problem discussed (Chapters 2, 3) is that of the role of scaffold proteins. The assembly of multiple signaling proteins into a complex by a scaffold protein guides many cellular decisions. Despite recent advances, the overarching principles that govern scaffold function are not well understood. We carried out a computational study using kinetic Monte Carlo simulations to understand how spatial localization of kinases on a scaffold may regulate signaling under different physiological conditions. Our studies identify regulatory properties of scaffold proteins that allow them to both amplify and attenuate incoming signals in different biological contexts. These properties are not caused by the well established prozone or combinatorial inhibition effect. These results bring coherence to seemingly paradoxical observations and suggest

that cells have evolved design rules that enable scaffold proteins to regulate widely disparate cellular functions.

After having studied how the amplitude of a signal can be regulated by a scaffold protein, we then focus our attention on how scaffold proteins regulate signaling dynamics (Chapter 3). These scaffolding proteins that direct the assembly of multiple kinases into a spatially localized signaling complex are often essential for the maintenance of an appropriate biological response. Although scaffolds are widely believed to have dramatic effects on the dynamics of signal propagation, the mechanisms that underlie these consequences are not well understood. Here, Monte Carlo simulations of a model kinase cascade are employed to investigate how the temporal characteristics of signaling cascades can be influenced by the presence of scaffold proteins. The simulations indicate that a major effect that scaffolds exert on the dynamics of cell signaling is to control how the activation of protein kinases is distributed over time. Scaffolds can influence the timing of kinase activation by allowing for kinases to become activated over a broad range of times, thus allowing for signaling at both early and late times. Scaffold concentrations that result in optimal signal amplitude also result in the broadest distributions of times over which kinases are activated. These calculations provide insights as to how the duration of a signal can be regulated in a scaffold mediated protein kinase cascade. Our results illustrate another complexity in the broad array of control properties that emerge from the physical effects of spatially localizing components of kinase cascades on scaffold proteins.

The observation that scaffold proteins regulate signal duration led us to consider mechanisms for how signal duration is detected (Chapter 4). Signal duration (i.e. the

time scales over which an active signaling intermediate persists) is a key regulator of biological decisions in myriad contexts such as cell growth, proliferation, and developmental lineage commitments. Accompanying differences in signal duration are numerous downstream biological processes that require multiple steps of biochemical regulation. Here, we present an analysis that investigates how simple biochemical motifs that involve multiple stages of regulation can be constructed to differentially process signals that persist at different time scales. We compute the dynamic gain within these networks and resulting power spectra to better understand how biochemical networks can integrate signals at different time scales. We identify topological features of these networks that allow for different frequency dependent signal processing properties. Our studies suggest design principles for why signal duration in connection with multiple steps of downstream regulation is a ubiquitous control motif in biochemical systems.

The notion that signal duration is such a widely exploited mechanism for mediating cell decisions, we then looked for other ways in which signal duration can be regulated. This led us to study the physics of multisite phosphorylation (Chapter 5). The utilization of multiple phosphorylation sites in regulating a biological response is ubiquitous in cell signaling. If each site contributes an additional, equivalent binding site, then one consequence of an increase in the number of phosphorylations may be to increase the probability that, upon dissociation, a ligand immediately rebinds to its receptor. How such effects may influence cell signaling systems is not well understood. A self consistent integral equation formalism for ligand rebinding, in conjunction with Monte Carlo simulations, is employed to further investigate the effects of multiple, equivalent binding sites on shaping biological responses. Multiple regimes that

characterize qualitatively different physics due to the differential prevalence of rebinding effects are predicted. Calculations suggest that when ligand rebinding contributes significantly to the dose response, a purely allovalent model can influence the binding curves nonlinearly. The model also predicts that ligand rebinding in itself appears insufficient to generate a highly cooperative biological response.

Furthermore, several applications to Immunology were also studied. Recent studies have reported that T cells can integrate signals between interrupted encounters with Antigen Presenting Cells (APCs) in such a way that the process of signal integration exhibits a form of memory. Here, we carry out a computational study using a simple mathematical model of T cell activation to investigate the ramifications of interrupted T cell-APC contacts on signal integration (Chapter 6). We consider several mechanisms of how signal integration at these time scales may be achieved and conclude that feedback control of immediate early gene products (IEGs) appears to be a highly plausible mechanism that allows for effective signal integration and cytokine production from multiple exposures to APCs. Analysis of these computer simulations provides an experimental roadmap involving several testable predictions.

In another study we investigated the effects of receptor clustering on the surface of a T cell (Chapter 7). T cell activation is predicated on the interaction between the T cell receptor and peptide-major histocompatibility (pMHC) ligands. The factors that determine the stimulatory potency of a pMHC molecule remain unclear. We describe results showing that a peptide exhibiting many hallmarks of a weak agonist stimulates T cells to proliferate more than the wild-type agonist ligand. An in silico approach suggested that the inability to form the central supramolecular activation cluster

(cSMAC) could underlie the increased proliferation. This conclusion was supported by experiments that showed that enhancing cSMAC formation reduced stimulatory capacity of the weak peptide. Our studies highlight the fact that a complex interplay of factors determines the quality of a T cell antigen.

Finally, we carry out calculations that seek to gain insight into signal transduction at a global scale across an entire proteome (i.e. set of proteins that constitute the entire signaling network). We present an analysis (Chapter 8) that attempts to gain insight into the mechanisms that govern growth factor derived mammalian signal transduction as probed using mass spectrometry based phosphoproteomics methods. We develop and apply techniques rooted in statistical mechanics and information theory to investigate both the global features and local interactions of these signaling networks. At the local scale, we infer the network of interacting tyrosine sites that governs signaling dynamics. We construct this interaction network between the phosphorylated sites by developing a model of network connectivity that maximizes the Shannon entropy[11] subject to the constraints of the pairwise correlations that are present in the data set. In developing this model, we predict the existence of few phosphotyrosine "hubs" that interact with many other tyrosine sites; these network hubs also lead to a small world network topology among the interacting phosphotyrosine sites. Furthermore, the model makes predictions of network connectivity between novel phosphotryosine sites identified by mass spectrometry and those of known nodes in the cognate signaling network.

Finally, ongoing work into the nature of signaling dynamics at the global scale is discussed. We first use a singular value decomposition to identify the slow characteristic modes of the signaling network that reveal the gross features of the underlying

biochemical network. We then briefly sketch ongoing work in the development of projection operator methods that aim to collapse all of the data onto these characteristic signaling modes by deriving generalized Langevin equations for describing the dynamics of these collective variables. In the future, these equations may serve as a framework for the predicting how these networks respond to perturbations.

Chapter 1.3. References to published work

Much of the work that is discussed in this thesis has been published elsewhere. Chapter 2, "Regulation of Signal Amplitude by Scaffold Proteins: Scaffold Proteins Confer Diverse Regulatory Properties to Protein Kinase Cascades" has been published in the Proceedings of the National Academy of Sciences[12]. Chapter 3, "Scaffold-mediated signaling dynamics: Regulation of signal duration and the statistical dynamics of kinase activation by scaffold proteins" is, at the time of writing, in press in the Public Library of Sciences – Computational Biology journal[13]. Chapter 4, "Signal Duration and the Time Scale Dependence of Signal Integration in Biochemical Pathways" is available as a preprint in the Cornell archives[14] and will be submitted to a peer reviewed journal after the time of this writing. Chapter 5, "Allovalency revisited: An analysis of multisite phosphorylation and substrate rebinding", has been published in the Journal of Chemical Physics[15]. Chapter 6, "T Cell Activation I: Computational Investigations into the Origins of Short Term Biochemical Memory in T Cell Activation", has been published in the Public Library of Sciences – ONE journal[16]. Chapter 7, "T Cell Activation II: Signal Integration at the Immunological Synapse and the Effects of Altered Peptide Ligands", is a collaborative effort among multiple experimental and theoretical scientists. The work mentioned in Chapter 7 consists of

parts of a larger paper that was published in the journal Immunity[17]. In chapter 7, I provide an overview of the project, mention the key results, and provide a detailed report that emphasizes the small part of the work that I contributed to the large collaborative effort. Chapter 8, "Signaling Networks at a Global Scale: Analysis of Phosphoproteomics Data Obtained from Mass Spectrometry", is work in progress and preliminary results are described. Finally, Chapter 9 is speculative and contains a discussion of future directions and continuations of work presented in this thesis.

Chapter 2 – Regulation of Signal Amplitude by Scaffold Proteins: Scaffold Proteins Confer Diverse Regulatory Properties to Protein Kinase Cascades

Chapter 2.1 Introduction

The sequential activation of multiple protein kinases constitutes a highly conserved intermediate step in eukaryotic cell signaling pathways, and is crucial for the regulation of numerous cellular decisions[18-20]. Common explanations for the ubiquity of these multi-leveled kinase cascades include the possibility that they amplify signaling and incorporate additional regulatory checkpoints that may improve the specificity and fidelity of the signal output[21-23]. In many instances (e.g. in several MAPK cascades), these kinase cascades are associated with scaffolding proteins that assemble multiple components of the signaling cascade in sequence (Fig. 2-1a)[24-27].

The general principles underlying how scaffold proteins function to influence signaling in protein kinase cascades are still poorly understood. Scaffold proteins are believed to be involved in many regulatory processes such as intracellular trafficking and pathway sequestering, and several factors have been shown to influence their signaling function[25]. For example, the relative concentration of scaffolding proteins has been shown to be a key variable that modulates signal output in many instances[28, 29]. Indeed, one signature of a scaffold protein is believed to be the appearance of a "bell-shaped" protein titration curve.

Recent reports also indicate that certain scaffolding proteins, such as Ste5 involved in the MAPK pathway of the Yeast mating response, can catalytically activate a MAPK upon binding by inducing autophosphorylation of the threonine residue in the TxY motif in the MAPK, Fus3[30]. More complexity is added by suggestions that some scaffolds may recruit phosphatases to their scaffold-bound substrates[25], or in contrast, protect scaffold bound kinases from phosphatase mediated deactivation[29, 31].

Although functions such as catalysis could be important for specific systems, the ubiquity of scaffolds suggests that the physical effects of tethering members of the cascade to a scaffold may have a functional role. It is difficult to ascertain that specific effects (e.g. catalysis or feedback) are absent in an experimental system. Therefore, it is problematic to study this potentially more generic function of scaffolds. Furthermore, a systematic variation of the many factors that may influence mechanisms through which spatial localization of kinases on a scaffold may affect signal propagation is currently not tractable. It would therefore be very useful to identify the most influential variables on which experiments should focus. For these reasons, we carried out computer simulations to study whether, how, and under what conditions, assembling a sequence of kinases on a scaffold affects signal propagation through a multi-tiered kinase cascade.

We investigated how scaffolds can influence protein motion, phosphorylation of downstream kinases by an active kinase, and phosphatase-mediated deactivation of kinases. The signaling module that we studied (Fig. 2-1) can be characterized by the following small number of parameters: the rate at which an active kinase can phosphorylate a downstream kinase, the rate at which phosphatases can remove a phosphate group from a kinase, the binding affinity of kinases to the scaffold or exchange rate, the relative concentration of scaffolds and kinases, and the parameters characterizing the mobility of the various proteins kinases.

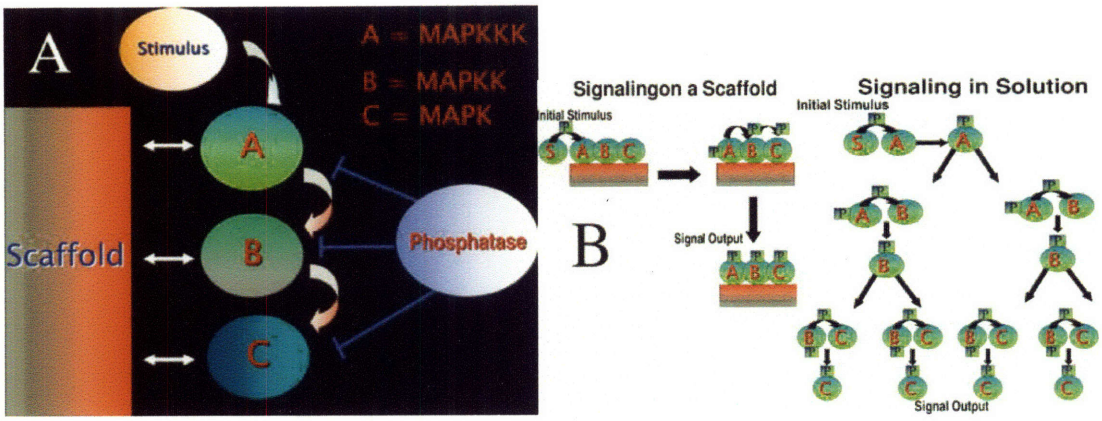


Figure 2-1. **Computer simulations model the effects of scaffolding a kinase cascade**

a.) In a model kinase cascade such as the MAPK cascade, an initial stimulus, S* (e.g. Ras-GTP), is recruited to and activates kinase A (MAPKKK). An active A (MAPKKK) in turn activates a B kinase (MAPKK) which then can activate kinase C (MAPK). Phosphatases are present than can encounter and deactivate activated kinases.

b.) Schematics are shown for the sequence of signaling events in solution and on a scaffold in our model. In order for a chemical reaction to occur in solution, the appropriate species must first come into contact with its substrate and then overcome a thermal energy barrier to model catalysis. When assembled on a scaffold, active kinases need only overcome the thermal energy barrier to activate its downstream target. Phosphatases are allowed to interact with active kinases are bound to the scaffold. Excluding phosphatases from interacting with scaffold proteins is also considered.

We studied how scaffolds influence signal propagation for different physiological conditions determined by values of these parameters. Schematics of how signaling occurs in solution and on a scaffold are shown in Fig. 2-1b. Other important issues that we examined using this model are the consequences of whether kinases bound to the

scaffold can phosphorylate downstream kinases that are in solution and whether phosphatases in solution can act on proteins bound to the scaffold.

The simple model we studied (with a relatively small number of parameters) allows us to meticulously study these different scenarios in depth. Yet, it provides enough relevant features to allow the mechanisms extracted from this model to be biologically meaningful. Results from our simulations indicate that, depending on cellular conditions, the spatial organization of kinases on scaffold proteins can either enhance or inhibit signal propagation through a kinase cascade. Specifically, scaffolds enhance signal propagation when prevailing conditions would lead to attenuation. Conversely, scaffolds inhibit the propagation of signals that would otherwise be greatly amplified. We discuss how these results provide a conceptual framework that sheds light on available observations and suggests further experiments that could help further elucidate the role of scaffolds in cell signaling processes.

Chapter 2.2 Model development and key variables

We simulated a model multi-leveled protein kinase cascade such as the Mitogen-Activated Protein Kinase Cascade (MAPK). An initial stimulus, S*, (e.g. Ras-GTP) can interact with a MAPKKK to activate it. The activated MAPKKK can then activate a MAPKK. The MAPKK, in turn, activates a MAPK. Finally, the MAPK activates a downstream species. The following sequence of signaling events,

$$\begin{aligned} MAPKKK &\xrightarrow{S^*} MAPKKK^* \\ MAPKK &\xrightarrow{MAPKKK^*} MAPKK^* \\ MAPK &\xrightarrow{MAPKK^*} MAPK^* \end{aligned} \qquad (2\text{-}1)$$

where an asterisk *, denotes a fully activated kinase, is also described in Fig. 2-1. Each of these above signaling events is modeled as an elementary chemical reaction. Each activated kinase can be deactivated by phosphatase-mediated catalysis. For notational convenience, MAPKKK will hereon be referred to as kinase A, MAPKK as kinase B, and the MAPK as kinase C.

Space is discretized on a 100 x 100 x 100 lattice (Supplementary Information, Fig. 2-7). If we assume a lattice spacing of 10nm, a typical diameter of a protein, the concentration of kinases in our simulation box is roughly 1μM for kinase A and kinase B and ~5μM for kinase C. In a physiological context, assuming the radius of the cell is about 10μm, this approximately corresponds to $\sim 10^5$ molecules of kinases A and B and a copy number of $\sim 5 \times 10^5$ for kinase C in our simulation. These relative numbers are commensurate with reported kinase concentrations[32]. 600 generic phosphatases are also present giving a 1:1 ratio of kinases to phosphatases. As described in the web supplement, in most cases, our qualitative results are not sensitive to variations of these numbers. For the results presented all events occur in the cytosol, but the consequences of restricting reactions to a membrane are also considered.

One important variable is the time it takes for an active kinase to encounter its substrate; our studies focused on experimentally relevant encounter times as diffusion constants[33] and aforementioned kinase concentrations were chosen accordingly (supplementary information). For most results, we consider the activation of protein kinases and phosphatase-mediated deactivation to occur through a single reactive collision involving reactants making contact and then overcoming a thermal energy

barrier. The results are qualitatively similar in many cases for catalytic mechanisms (supplementary information, Fig 2-8).

The model as defined above involves the following biophysical parameters: E, the energy barrier for disassociation of a bound species from a scaffold, E_2 the energy barrier for association of a kinase to a scaffold, E_3, the energy barrier for activation of a kinase by an appropriate target enzyme, E_4 (Energy barrier for deactivation of a kinase by a neighboring phosphatase). The amplitude of the initial stimulus is an important quantity in cell signaling. We define a variable, $\sigma \equiv \frac{[S^*]}{[A]_0}$ (the amplitude of initial stimulus scaled to a characteristic density of protein kinase, $[A]_0$). Other important variables are: $\lambda \equiv \frac{P_{Diffusion}}{P_{Reaction}}$ (the ratio of probabilities for attempting diffusion and reaction moves in the simulations) which is a measure of protein mobility and $\xi \equiv \frac{[Scaffold]}{[A]_0}$ which is the ratio of scaffold density to a characteristic density of kinases.

In order to quantify how signals propagate along the cascade, we define an amplification factor, ϕ, that measures the fractional change in activated signaling species as the signal propagates through the cascade (i.e. the ratio of the numbers of an activated downstream kinase at the end of the cascade relative to that of the first kinase); $\phi \equiv \left\langle \frac{C^*}{A^*} - 1 \right\rangle$, where * denotes a fully activated species, and the brackets indicate an average over many simulations (analogous to a population average over many cells).

ϕ is positive if the signal is amplified as it propagates along the cascade, and is negative if it is attenuated. The absolute magnitude of the signal, $\theta \equiv \left\langle \frac{C^*}{[A]_0} \right\rangle$, is also considered.

Sensitivity amplification[34], $\left\langle \frac{d(\ln C^*)}{d(\ln S^*)} \right\rangle$, another metric often used for studying signaling cascades and their input-output characteristics, is not considered here since our study focuses mainly on the propagation of a signal's amplitude – how scaffold proteins could shape the dose response (i.e. input-output characteristics) of a signaling system has been previously studied theoretically [24, 29, 31]. It has been suggest that for cases when enzymes are saturated in solution, scaffolds would decrease the sharpness of the dose-response curves and can convert switch-like to graded responses.

We compute the quantities of interest using a kinetic Monte Carlo algorithm[35, 36] that is described in the methods section and supplementary information. For reasons mentioned therein, a signaling model using ordinary differential equations (ODEs) was not used.

Chapter 2.3 Results

Chapter 2.3.1 Scaffold proteins can amplify signals that are attenuated in the absence of a scaffold

We first considered the situation where basal phosphatase activity is high, making the cascade intrinsically difficult to activate. In this circumstance, weak signals are rapidly attenuated and do not propagate regardless of whether or not a scaffold is present (data not shown). Also in the absence of scaffolds strong signals ($\sigma \sim 1$), are attenuated

($\phi < 0$). We obtained this result by taking the affinity of kinases to the scaffold to be sufficiently weak so that, on average, these proteins are not bound to the scaffold.

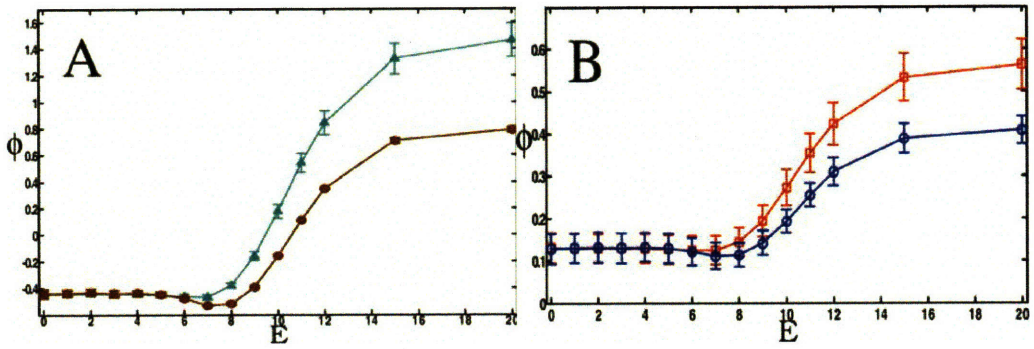


Figure 2-2. **Scaffold proteins can amplify signals that would otherwise attenuate**

The case of high basal phosphatase activity, $E_4 = 0$. Shown below are calculated values of signal amplification, φ, $\phi \equiv \left\langle \frac{C^*}{A^*} - 1 \right\rangle$ (a.) and signal magnitude, θ ($\theta \equiv \left\langle \frac{C^*}{[A]_0} \right\rangle$) (b.), for increasing values of the kinase-scaffold binding affinity E. The A (MAPKKK) and B (MAPKK) concentrations equal that of the scaffold, while the concentration of C (MAPK) is five times larger (other situations are described in the supplementary information). A strong stimulus, σ = 1 ($\sigma \equiv \frac{[S^*]}{[A]_0}$), is used. In each panel, two cases are considered: a "constrained" case where species bound to the scaffold can not activate species in solution (diamonds in (a) and circles in (b)) and an "unconstrained" case where species bound to the scaffold can activate species in solution (triangles in (a) and squares in (b)).

When the affinity of the kinases to the scaffold increases, kinases begin to assemble onto the scaffold. When the affinity is sufficiently strong, available kinases are

bound to their corresponding binding sites on the scaffold and this scenario can be interpreted as having a scaffold present. We find that, when phosphatase activity is high, and the initiating stimulus is strong, assembling kinases onto the scaffold results in a sharp increase in the number of active kinases as the signal propagates along the cascade (Fig. 2-2a). In this circumstance, assembling kinases on a scaffold allows for significant signal amplification. For reasonable parameter values, a ~100% increase in signal output ($\varphi \sim 1$), with respect to the first member of the cascade, can result. Amplification is also evident in the absolute magnitude of the signal output (Fig. 2-2b). Thus, when phosphatase activity is sufficiently high, a scaffold allows for the effective propagation of a signal that would otherwise be significantly quenched. Note that this amplification of signal propagation occurs even when phosphatases can act on scaffold-based kinases.

Chapter 2.3.2 Scaffold proteins can suppress signals that are amplified in the absence of a scaffold

Now we consider situations in which basal phosphatase activity is low, rendering the cascade easy to activate. In this circumstance, strong ($\sigma \sim 1$) and weak ($\sigma << 1$) signals are strongly amplified in the absence of a scaffold as there is little to impede the propagation of signals (Figs. 2-3a,b). The magnitude of ϕ is much larger compared to the situation where phosphatase activity is high because, in this case, the signal is amplified exponentially as each activated kinase can activate many downstream targets.

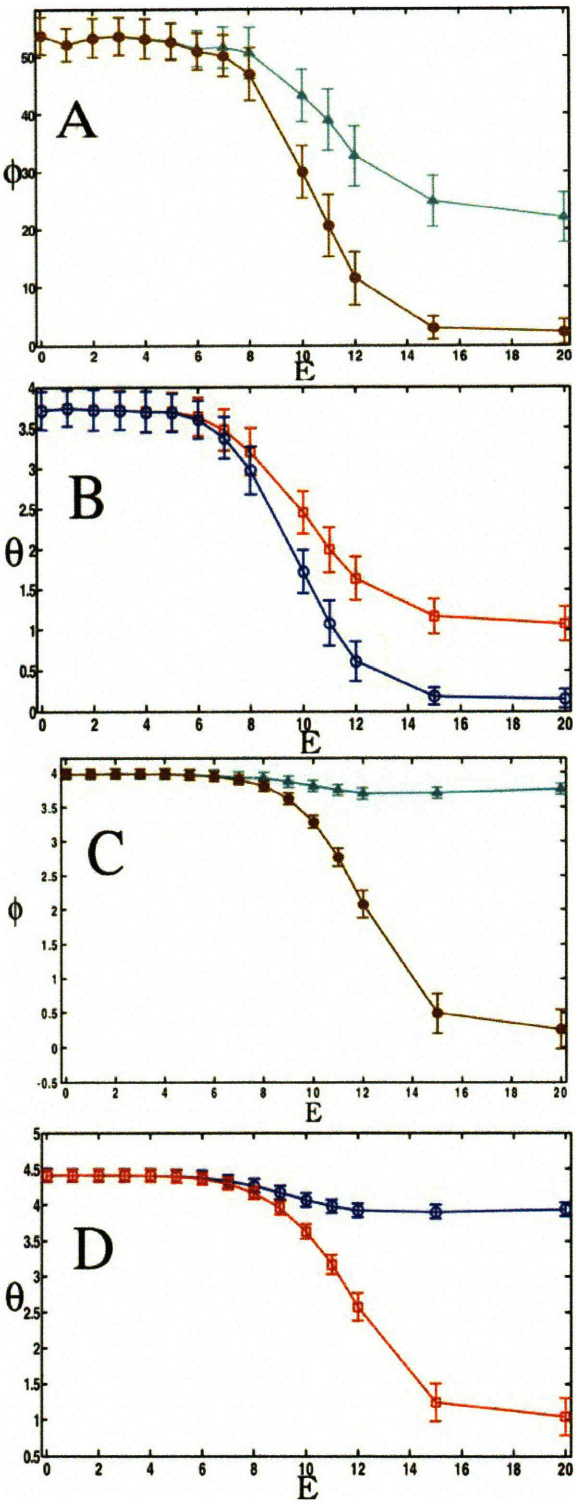


Figure 2-3. **Scaffold proteins attenuate signals that would otherwise strongly amplify.**

Low basal phosphatases levels are considered (E_4 = 6). Signal amplification φ, ($\phi \equiv \left\langle \frac{C^*}{A^*} - 1 \right\rangle$) (a,c), and signal magnitude θ, ($\theta \equiv \left\langle \frac{C^*}{[A]_0} \right\rangle$), (b,d) are considered as a function of scaffold binding affinity, E. Two cases are shown: (a,b) a weak stimulus (σ << 1, $\sigma \equiv \frac{[S^*]}{[A]_0}$) of small amplitude and low basal phosphatases levels(E_4 = 6). (c,d) a stimulus of large amplitude (σ = 1, $\sigma \equiv \frac{[S^*]}{[A]_0}$); all other conditions are the same as those reported in Figure 2. As shown in these plots assembling signaling components onto a scaffold by increasing E results in significantly lower amplification and signal output. Again, in each panel, two cases are considered: a "constrained" case where species bound to the scaffold can not activate species in solution (diamonds in (a,c) and circles in (b,d)) and an "unconstrained" case where species bound to the scaffold can activate species in solution (triangles in (a,c) and squares in (b,d)).

In striking contrast to the situation where phosphatase levels are high, for low phosphatase activity, assembling kinases on a scaffold greatly suppresses signaling and limits signal amplification (Figs. 2-3a-d). For weak signals, this effect is accentuated if scaffold-bound kinases can only interact with downstream kinases that are also attached to the scaffold (Figs. 2-3a-d). Although signal amplification still occurs, our simulations demonstrate that the extent to which amplification occurs is severely limited (and approaches nonexistence) when scaffold-bound kinases are constrained to interact only with other scaffold-bound kinases.

If scaffold-bound kinases are allowed to activate their downstream targets that are not bound to the scaffold, scaffolds have only a small effect on the propagation of strong

signals when phosphatase activity is low (Figs. 2-3b,d). The slight dampening of the signal in this case is largely due to a smaller frequency of collisions between proteins bound to the scaffold and those in solution compared to that of mobile proteins in solution.

The ability of scaffolds to limit amplification can also be manifested in other interesting ways. Our calculations show that, provided phosphatase activity is sufficiently low and scaffold-bound kinases cannot activate free kinases, increasing the exchange rate (i.e. lowering the energy barrier for disassociation) characterizing the binding of the last kinase (kinase C) to the scaffold can result in more efficient signal propagation (Fig. 2-4). This result suggests that the observed[37] low affinity of Erk to KSR (a scaffold associated with a MAPK module in mammalian cells) may serve to enhance signal propagation. It is also tempting to suggest that this could be one (of many) factor underlying the recent observation that a mutation, that blocks binding of the MAPK, Fus3, to the scaffold, Ste5, but keeps the docking site to the MAPKK, Ste7, intact, results in greater signaling output in the yeast mating response[30] as this could effectively increase the exchange rate.

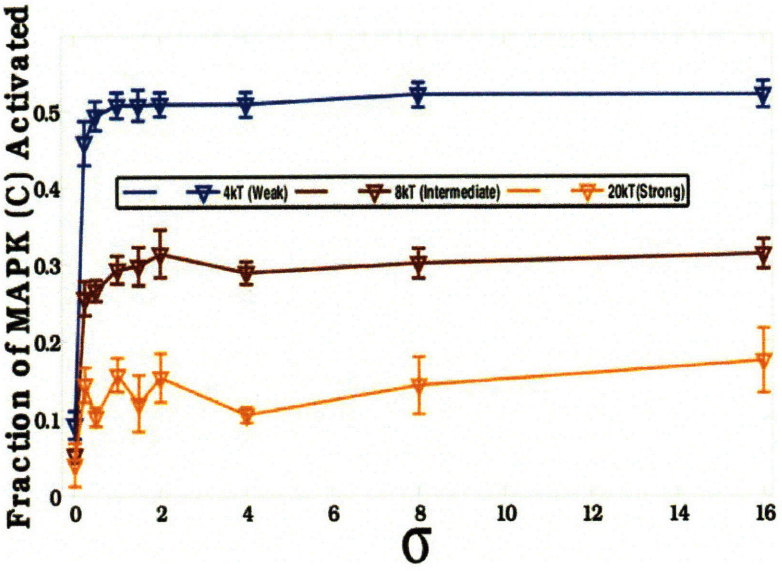


Figure 2-4. **Serial engagement of the last kinase can greatly influence signal output in a MAPK cascade.**

We consider the scenario where the first two kinases, A and B (MAP3K, MAP2K) bind tightly ($E_2 = 20$) and the binding affinity C (MAPK) is varied as follows: weak ($E_2 = 4$), intermediate ($E_2 = 8$), and strong ($E_2 = 20$) disassociation energies. The ordinate represents the fraction of activated C (MAPK) proteins (a measure of signal output) and the abscissa is the scaled strength of signal, σ, $\sigma \equiv \frac{[S^*]}{[A]_0}$. The signal output is largest when kinase C (MAPK) is allowed to rapidly disassociate ($E_2 = 4$) from the scaffold. Data presented are for a 1:1:10:1 ratio of A:B:C:Scaffolds.

Chapter 2.3.3 Mechanism underlying why scaffolds can both amplify and limit signal propagation

A unified conceptual framework explaining why scaffolds can both enhance and attenuate signaling emerges from our study. Suppose a kinase such as kinase A in Fig. 1a, is activated. In the absence of a scaffold, the number of molecules, N, of the downstream kinase (kinase B) that it could potentially phosphorylate must be enclosed in

a volume that grows with the diffusion coefficient of the upstream kinase (kinase A) and the time (τ_p) it takes phosphatases to deactivate it. Scaffolds impose a stoichiometric constraint that limits the number of molecules (N_{scaf}) of the inactive downstream target (kinase B), that can potentially be phosphorylated, to those that are bound to the scaffold. When $N > N_{scaf}$, an activated kinase (kinase A) can phosphorylate more of its substrates (kinase B) in the absence of the scaffold. This is the situation when phosphatase activity is low (τ_p is large), and so a scaffold limits signal amplification (Fig. 2-3). However, when phosphatase activity is high, N can be less than N_{scaf}; i.e., signal propagation in the absence of a scaffold is limited by the short time available for encounters between kinases before phosphatases inactivate them. Now, scaffolding results in amplification (Fig. 2) because the spatial proximity of kinases on a scaffold reduces the encounter time, and this effect dominates over the inhibitory stoichiometric constraint.

More precisely, N, must be contained in the volume, $\Omega \sim (D\tau_p)^{d/2}$, where d is the spatial dimension (d =2 for a cell membrane, and d =3 for signaling in the cytosol or on an endosome). Then,

$$N = \int_{\Omega} d^d x \; \rho_B \qquad (2\text{-}2)$$

where ρ_B is the concentration of inactive B kinases. If proteins are assembled onto a scaffold, the only molecules of the inactive B that can potentially be phosphorylated by an active A molecule are those that are on the scaffold (i.e. they are enclosed within a volume,

$$\Omega \sim (L_{scaf})^d, \qquad (2\text{-}3)$$

where, L_{scaf}, is the distance between kinase B and kinase A on the scaffold. The number of kinase B that could be potentially activated is then

$$N_{scaf} = \int_{\Omega} d^d x \left[\sum_{i=1}^{i=N_{scaf}} \delta\left(\vec{x} - \vec{x}_i\right) \right], \qquad (2\text{-}4)$$

where $\delta\left(\vec{x}\right)$ is the Dirac delta function and N_{scaf} is the number of kinases that can be activated on the scaffold (1 in our case). When kinases in solution can exchange with scaffold bound kinases, the effective N_{scaf} in the sum can be larger. When the system is characterized by conditions such that $N > N_{scaf}$, an activated A can phosphorylate more B proteins in the absence of the scaffold because the stoichiometric constraint imposed by scaffolding is inhibitory. This is the situation when phosphatase activity is low (τ_p, Ω, and N are large), and so a scaffold limits signal amplification (Fig. 2-3). However, when phosphatase activity is high, N can be less than N_{scaf} because τ_p and Ω are small; i.e., signal propagation in the absence of a scaffold is limited by the short time available for encounters between kinases before phosphatases inactivate them. In these circumstances, scaffolding results in amplification (Fig. 2-2).

Also, under these conditions, significant signal amplification occurs even when there are equimolar concentrations of signaling components bound to the scaffold. The time it takes for a scaffold-bound downstream kinase to be activated by a scaffold-bound upstream kinase, is much shorter than the time required for the stimulus, S*, to be recruited to the complex and activate the first kinase in the sequence. Therefore, when phosphatases render kinases inactive on the scaffold, the kinases further down the cascade become reactivated on average much more quickly than those at the beginning of

the cascade. This causes scaffolds to influence kinase activation in a hierarchical fashion. If a scaffold bound B* or C* is deactivated, a scaffold bound A* or B* quickly reactivates B or C. However, if A* is deactivated, a much longer time is required for a reactive collision with S*. One then notices a positive gradient of active signaling molecules when moving along the cascade. Our simulations suggest that this effect can potentially be very significant – allowing for many-fold changes in signal output with reasonable parameter values. This effect can be seen in the positive values of φ; in Fig 2-2a for example, $\varphi \sim 1$ implying a ~100% increase in active molecules of C as compared to A.

Chapter 2.3.4 The network of interactions between phosphatases and their target kinases can influence the function of scaffold proteins

Although phosphatases are in general believed to play a crucial role in regulating kinases cascades, the precise ways in which phosphatases interact with members of kinase cascades in different contexts are not fully understood and are currently the subject of intense study[38, 39]. Some have also suggested that, due to steric constraints, phosphatases may not be able to interact with activated kinases that are bound to a scaffold [29, 31]. In these theoretical studies, phosphatases were prevented from acting on scaffold-bound kinases and this enhanced signaling.

We have studied the consequences of different scenarios by which phosphatases can interact with scaffolded kinase cascades (appendix, Fig. 2-9). We find that when individual kinases are protected from deactivation when bound to a scaffold, qualitative

behavior in the signal amplitude does not change (Figs 2-9a,b). Similar results are obtained when all three kinases are protected from deactivation when bound to the scaffold. Alternatively, if certain kinases, upon activation, are never deactivated by phosphatases, then some qualitative findings can change as discussed in the supplementary information.

For instance, in Fig. 2-9c, when phosphatases can not deactivate kinase C, assembling kinases on the scaffold with a high affinity reduces signal output; this is because some kinases in solution do not become activated during the course of signaling since they are never able to occupy a complete signaling complex. This effect is removed when kinases that are bound to the scaffold can interact with their substrates in solution. For these cases (Fig. 2-9c), such perturbations result in scaffolds having little to no effect on the signal output.

Chapter 2.3.5 Importance of scaffold concentration

One signature of a scaffold protein is believed to be the presence of non monotonic behavior ("bell shape") in a scaffold protein titration curve. If scaffolds are essential for activation, then too few scaffolds results in reduced output; alternatively, at high concentrations, scaffolds also would inhibit signaling since molecules that are activated in sequence are less likely to be bound to the same scaffold (i.e. the prozone effect)[24]. So, we considered how our results depend on scaffold concentration under conditions of high (Fig. 2-9a) and low (Fig. 2-9b) basal phosphatase activity. We find a "bell-shaped" curve for high phosphatase activity (Fig. 2-9a) when "titrating" along the relative scaffold concentration, ζ. For conditions of low phosphatase activity (Fig. 2-9b), the non monotonic behavior in signal output that appears in Fig. 2-9a is no longer

present. This is because, in this case, both the low phosphatase, inhibitory regime that characterizes the scaffold-mediated signaling as well as the prozone effect are at work.

Finally, many parameters governing the behavior of cell signaling pathways, such as diffusion coefficients, binding affinities, and catalytic rate constants, are unknown and are likely to greatly differ in individual physiological contexts. It is therefore essential that we understand how our results depend on unknown parameter values. We performed extensive sensitivity analysis with respect to all relevant parameters to examine the robustness of the qualitative behavior to parameter values. Our results demonstrate that in most cases (supplementary information), the qualitative function of assembling kinases onto a scaffold described above is insensitive to variations over wide ranges of these parameters.

Chapter 2.3.6 Influence of the network of interactions between phosphatases and their target kinases

We investigated two distinct hypotheses. In Figs. 2-9a and 2-9b, results are shown for cases in which a particular kinase can be inactivated in solution by a phosphatase but is protected from such deactivation when bound to the scaffold. Figs. 2-9c and 2-9d, show results from simulations in which a particular kinase is protected from phosphatase mediated deactivation (i.e. the rate of inactivation is zero) regardless of whether it is in solution or bound to a scaffold as could be the case in a mutant kinase -- no spontaneous activation occurs on the timescale of our simulation since the energy required to remove a phosphate group in the absence of an enzyme is very large in biological contexts.

We calculated θ for situations where scaffolds enhance (Figs. 2-9a,c) and where they inhibit (Figs. 2-9b,d) signal amplification in the "wild type" case (Figs. 2-2 and 2-3). All other parameters are the same as those used in Figs. 2-2a,b and Figs. 2-3c,d, and scaffold bound kinases can not act on their targets that remain in solution.

We first investigate the situation where kinases are protected from phosphatases only when they are bound to the scaffold. As seen in Fig. 2-9a,b qualitative results are identical to their analogous cases in Figs. 2-2, 2-3. Quantitative signal output does however change. This is because the value of N_{scaf}, and hence signal output, increases since some kinases are permanently active when bound to the scaffold.

Now, we consider the situation in which phosphatases do not act on a particular kinase irrespective of whether or not it is bound to the scaffold or in solution. First, we discuss the case of high constitutive phosphatase activity (Fig. 2-9c). For the case where phosphatases can not act upon the first kinase (A), we find that the qualitative behavior in signal amplitude is the same as that in Fig. 2b (Fig. 2-9c diamonds) and scaffolds amplify signals. However, when phosphatases do not act on kinase B (Fig. 2-9c circles), scaffolding makes only a minor difference for the case of high phosphatase activity. In solution, when the first active kinase (A*) is protected from phosphatase action, the step involving C→C * is still hindered by the quick deactivation of B* by phosphatases. Our results imply that the short encounter time for this step on a scaffold allows it to amplify signals in this circumstance. However, when kinase B is protected from phosphatase activity, the step involving C→ C* is not hindered by the inactivation of B* and the advantage of having a short encounter time becomes less important.

This suggests something counterintuitive; scaffold proteins could potentially inhibit signaling if the final (kinase C) kinase along the cascade is protected from phosphatase interactions. This is observed in Fig. 2-9c (crosses) where kinase C is protected. This effect follows from a consideration of the excess kinases in solution that are never allowed to assemble onto a scaffold since the affinity of a kinase to a scaffold is sufficiently strong so that the kinases do not exchange from the scaffold in the timescale of the simulation. However, such an inhibitory function is eliminated when downstream kinases are allowed to exchange from the scaffold (in reasonable times) as all kinases eventually become permanently activated.

Finally, for low basal phosphatases levels, similar qualitative behavior to that observed in Fig. 2-3 is seen at low phosphatase activities. For this case, in both the wild type and kinase-protected scenarios (Fig. 2-9d), the inhibitory effects of the scaffold dominate provided that scaffold bound kinases can not interact with their substrates in solution and the exchange rate from the scaffold is not too fast.

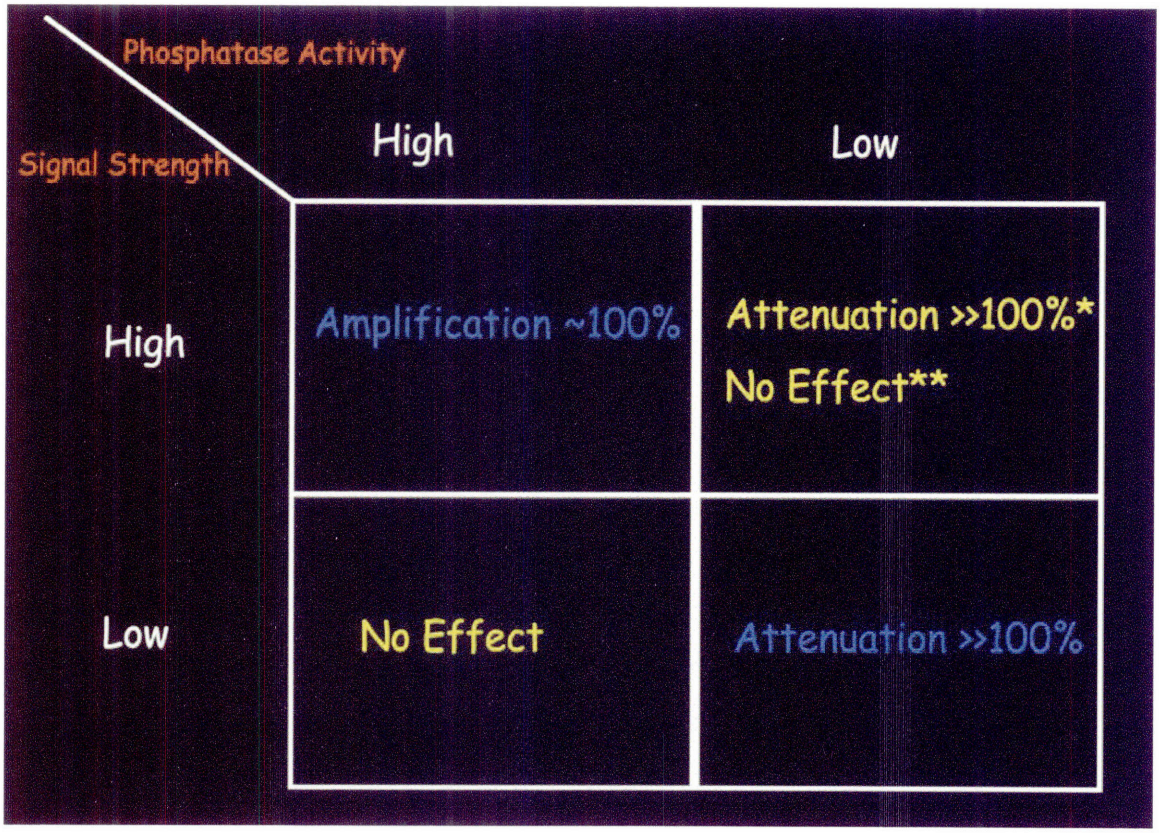


Figure 5. **Summary of the different regimes that characterize scaffold-mediated signal transduction**

The characterization of four regimes of scaffold-mediated signal transduction: (1) high signal strength and high phosphatase activity, (2) high signal strength and low phosphatase activity, (3) low signal strength and high phosphatase activity, and (4) low signal strength and low phosphatase activity. * indicates the case when kinases bound to a scaffold can not phosphorylate their downstream substrates that remain in solution. ** indicates the case when kinases can interact with their downstream targets that are present in the solution. The absence of * or ** indicates that the effect occurs regardless of whether or not scaffold bound proteins can activate downstream targets in solution. These results summarize our findings when phosphatases are allowed to act on scaffold-bound kinases, and reflect the balance between stoichiometric constraints and removal of transport limitations by scaffolds. As described in the text and Fig. S4, preventing

phosphatases from acting on scaffold-bound kinases does not alter these results substantially.

Chapter 2.4 Discussion

A unified conceptual framework that explains why scaffolds can both enhance and attenuate incoming signals, and why this is biologically important, emerges from our findings; a summary of the key findings is given in Fig. 2-5. Our results suggest that a robust system that is resistant to spurious noise but needs to respond vigorously to a strong stimulus should be designed with high basal phosphatase activity and scaffolds. An example is provided by T cells where discriminating between minute amounts of antigenic stimulus and spurious noise is critical. Basal phosphatase activity is high in T cells (phosphatases, PP5a , PP2, among others have been shown to interact with MAPK components in mammalian cells[38, 39]), and the KSR scaffold[40, 41] is known to amplify signaling through the MAPK pathway as measured by ERK activation.

Conversely, a system that needs to be highly sensitive to stimuli should be designed such that the cascade is intrinsically easy to activate (e.g., low phosphatase activity). Such a system would be overwhelmed, however, if exposed to a strong stimulus. Scaffolding kinase components can limit signal amplification and prevent the potentially catastrophic consequences of amplifying strong stimuli in this case. Signal integration during phototransduction requires a response to a few photons, and this is facilitated by a cascade that is easy to activate. Scaffolding such a cascade using INad[42] inhibits signal amplification in drosophila.

Our results suggest that scaffold proteins can intrinsically amplify strong signals that would otherwise not propagate efficiently, but in other instances, they can insulate kinase cascades from propagating signals that would otherwise be strongly amplified

with potentially deleterious consequences. As signaling often results in the upregulation of phosphatase expression, it is interesting to speculate that the role of scaffolds may change with time; scaffolds could take on positive and negative regulatory functions at different time points in the course of signal transduction. By amplifying attenuating signals when phosphatases are upregulated and abating amplifying signals when phosphatases are downregulated, signaling specificity and fidelity, to some extent, could then be obtained by having the scaffold provide the right balance of positive and negative control of the signal amplitude. Scaffold proteins would then confer additional, intrinsic, feedback mechanisms to an already carefully regulated, highly evolved, complex cellular process. The diverse consequences of spatially organizing kinases on a scaffold revealed by our studies would be further modulated by processes such as catalytic modifications and feedback control.

Our results highlight the dual positive and negative regulatory properties that scaffold proteins confer to kinase cascades. Such properties are predicted to have consequences on shaping the input-output characteristics of the signaling cascade. Since scaffolds limit signal amplification when phosphatase levels are low, the model predicts that, under these conditions, dose-response curves would appear less sharp in the presence of scaffolds. This is because each stimulating molecule has the ability to activate many more than one downstream target in solution whereas the scaffold limits this number. Conversely, under conditions in which signals attenuate in solution (high phosphatase levels), scaffolds increase the sharpness of the dose-response curve only slightly. Such scenarios presuppose that the kinases are not operating in saturating or "zero order ultrasensitive" conditions in solution. The latter situations would allow

scaffolds to change the dose response from an "all or none" to a graded response as has been suggested before (7, 12, 14).

We hypothesized that scaffold-mediated signaling could be influenced by the rate at which kinase C (the MAPK) exchanges with the scaffold as this would permit a scaffolded signaling complex to generate many activated C species. Therefore, we investigated the effects of the affinity of the third member of the cascade to the scaffold. We found that in some situations, decreasing the binding affinity of kinase C to the scaffold can result in a larger response. This follows from the ability of each signaling complex to potentially activate many of its downstream substrates when the substrate can quickly exchange with the scaffold. We noted that this prediction could explain the observed low affinity of ERK to the KSR scaffold [37]. We also speculated that this may underlie recent data that demonstrates that mutating the docking site of Fus3 to the scaffold Ste5 results in a larger signal output since it reduces the strength of interaction of this kinase with the scaffold[30]. However such simple explanations for these observations may be confounded by specific features of importance to particular signaling mechanisms. Yet, in light of our results, it is tempting to suggest that the affinity between the MAPK and the scaffold may be a crucial parameter that has been evolutionarily fine-tuned for specific systems.

We also investigated how the topology of the kinase/phosphatase interaction network affects signal output. When kinases are protected from phosphatases only when bound to the scaffold, qualitative results (Fig. 2-9a,b) are the same as when phosphatases can act on scaffolded kinases (Figs. 2-2, 2-3). But, in some instances (Fig. 2-9c), changes to the network topology (i.e., preventing phosphatase action on certain members of the

kinase cascade regardless of whether or not they are attached to the scaffold) can change the functional role of assembling kinases onto a scaffold.

Finally, we studied how our results depend on scaffold concentration. For high basal phosphatase activity, we found a "bell-shaped" titration curve. This is consistent with previous work (12) which, however, did not consider the potentially inhibitory role of scaffolds revealed by our studies of situations where basal phosphatase activity is low (Fig. 3). When scaffolds limit signal propagation, no such non-monotonic behavior was seen (Fig. 2-10b). Instead, signal output decreases monotonically with increasing scaffold concentration.

Our computational studies suggest general molecular mechanisms for how scaffold proteins can regulate signal transduction in diverse ways. Thus, our results provide possible clues for why scaffolds are involved in so many critical cellular pathways and how their improper regulation can lead to disastrous phenotypic outcomes[43-45]. We hope that our attempt to elucidate some general principles that govern scaffold function provides conceptual guidelines for future experimentation.

Chapter 2.5 Methods

Chapter 2.5.1 Simulation methodology

We simulate a model protein kinase cascade such as the mitogen-activated protein kinase (MAPK) cascade (Fig. 1a) in the presence and absence of a scaffold with a Metropolis-based kinetic Monte Carlo algorithm[35, 36]. The Monte Carlo simulations (Fig. 2-1b) allow us to monitor the relevant stochastic processes that occur in cell signaling. Proteins are represented as discrete objects, occupying a site on a lattice of dimensions 100 x 100 x 100. Reflecting (i.e. hard-wall) boundary conditions exist at

each of the faces of the cubic lattice. The system is not periodically replicated. Proteins can diffuse (i.e. translate on the lattice in random directions), bind and unbind, and undergo state transformations according to a prescribed reaction network (Fig. 2-1a,b). Protein motion is subject to excluded volume (steric) constraints in that no two proteins can occupy the same site on the lattice. Chemical (state) transformations and binding events are modeled as thermally activated processes with associated energy barriers, E_i, for activation, inactivation, binding and unbinding reactions.

In a Monte Carlo step, n trials are attempted, where n is the number of proteins in the simulation. For a given trial, a protein is first chosen at random with uniform probability. A displacement move in a particular direction is then attempted with probability, $P(diffusion) = \frac{1}{6} P_{diffusion} \min\{1, \exp(-E_\infty)\}$ where $P_{diffusion}$ is the probability of attempting a diffusion move. Excluded volume is accounted for by imposition of an infinite energy barrier, E_∞, located at sites containing other proteins but is zero everywhere else. Upon considering all possible nearest neighbor interactions, reaction moves, as determined by the network topology, are tried with probability, $P(reaction) = P_{reaction} \min\{1, \exp(-E_i)\}$ where $P_{reaction}$ is the probability of attempting a reaction move, E_i is the energy barrier for the i^{th} reaction scaled with respect to k_BT (Boltzman's thermal energy) and the Kronecker delta function imposes the constraint that only nearest neighbors can interact.

Chapter 2.5.2 Parameter choices

For the results presented in the main text, we simulated a system consisting of 200 A (MAPKKK) species, 200 B species (MAPKK), and 1000 C species (MAPK) and 600

Generic Phosphatases. Scaffold proteins are modeled as structural elements and are represented as rigid, immobile, objects with each unit occupying one lattice site. Scaffolds do not change the activation energies for phosphorylation reactions and catalytic steps when species are bound to the scaffold. Each unit contains a binding site that is specific to a particular type of kinase and occupies one site on the lattice. We allow both inactive and active kinases to potentially bind to their specified binding sites and catalysis to take place when two appropriate species come into contact.

As stated in the main text, the reaction network was modeled in the simplest possibly way by considering each elementary reaction event to comprise of a single reactive, thermally-active collision. The reactions used in the main results are:

$$\begin{aligned} A + S^* &\rightarrow A^* + S^* \\ B + A^* &\rightarrow B^* + A^* \\ C + B^* &\rightarrow C^* + B^* \\ A^* + P &\rightarrow A + P \\ B^* + P &\rightarrow B + P \\ C^* + P &\rightarrow C + P \end{aligned} \qquad (2\text{-}5)$$

Each reaction occurs in solution and when bound to the scaffold. Each species also can bind and unbind from the scaffold allowing for 27 possible chemical states of signaling complexes.

The energy barrier for association, $E_2 = 0$, implying that proteins will bind to the scaffold when in contact with their selective binding sites. E, the energy barrier for disassociation from the scaffold, can range from 0 to 20. Note, for $E = 0$, statistically no proteins are bound to the scaffold, and for $E = 20$, each binding site on a scaffold is occupied (on average) provided enough kinases are present. For simplicity, we consider the association and disassociation constants for each site on the scaffold to be the same

unless otherwise noted. The energy barrier for activating a species by the appropriate kinase, E_3, is taken to be zero, implying that phosphorylation will occur when an activated kinase collides with its appropriate substrate. E_4, the energy barrier for deactivating a kinase, varies in the simulation trials and can take on values ranging from 0 to 10. For high phosphatase activity (i.e. a cascade that is intrinsically difficult to activate) , E_4= 0. For low phosphatase activity (i.e. a cascade that is intrinsically easy to activate), E_4= 6. It is important to note that varying the number of phosphatases would have similar effects as varying E_4 . In this regard, E_4 then sets a basal level of phophatase activity. For the case of a low signal, σ is taken to be 0.01. For the case of a high signal, σ is taken to be 1.0 or 0.25 as specified. ξ is set at 1, corresponding to one stoichiometric equivalent with respect to the first kinase in the cascade. Unless noted otherwise, all parameters used in particular simulations are those given above. But, we also studied the sensitivity of our results to variations in parameters.

Chapter 2.5.3 Signaling in the cytosol vs. a membrane

We considered all events to occur in the cytosol (i.e. in three dimensions). However, it has been shown that the MAPK cascade can become active in the course of membrane-proximal signaling events with a scaffold protein, such as KSR, being recruited to the plasma membrane. Also in some cases, scaffold mediated signaling may be confined to the membranes of endosomes. Considering the signaling cascade on a plasma membrane does not change the qualitative results obtained from the simulations. The quantitative behavior of the signaling cascade, however, changes.

The effects that we report (i.e. amplification vs. attenuation) become, in some cases, more pronounced when signal transduction occurs on a membrane such as the plasma or nuclear membrane. Differences in the quantitative behavior follows from the following effects: the mobility of proteins in the plasma membrane is significantly smaller than for proteins in the cytosol, and so the area that an active kinase can sweep out becomes smaller in two dimensions; also, the time it takes for a phosphatase to deactivate a kinase changes accordingly. A very interesting recent study reports some of these effects in a scaffolded mammalian MAPK system[46]. Their study finds that the sensitivity of the MAPK pathway greatly increases when signaling is confined on a membrane.

Chapter 2.5.4 Importance of protein diffusion

An important variable that determines the role of scaffolding a kinase cascade is the amount of time required (τ_{ec}) for an active kinase to encounter its downstream target. For simple diffusion, in three dimensions, $\tau_{ec} \sim \frac{1}{DC^{2/3}}$ where D is the diffusion constant and C is a typical concentration of kinases. Experiments indicate that τ_{ec} is significantly larger than 1μs (on the order of 10^{-4}s - 10^{0} s)[33]. Our studies focused on these experimentally relevant conditions.

Monte Carlo time can be related to real time by equating a diffusion constant of a typical protein in the cytoplasm (i.e. 10μm^2/s)[33] with the diffusion constant used in most of our simulations (1 (lattice spacings)^2)/mcstep), and taking a value of the lattice spacing to be 10nm. The cell is considered to have roughly a 6μm radius. This leads to a conversion factor, 1 mcstep ~ 1 μs. Typical physiological values of τ_{ec} range from

roughly 10^{-4} s – 1 s. These values of are obtained from consideration of typical values of D (for a ~50kD protein) in the cytoplasm($10\mu m^2/s$)as well as upper and lower bounds of kinase concentrations[32, 47]. Fig. 7 gives an example of how protein mobility can influence the role of scaffold mediated signal transduction.

Chapter 2.5.5 Variations in kinases concentration and catalytic rates and mechanisms

Fig. 7 illustrates some of the dependence of the results on kinase numbers and catalytic rates. We also explored the effects of incorporating enzymatic mechanisms for kinase activation rather than single reactive collisions. Specifically we studied the following mechanism:

$$E + S \underset{k-1}{\overset{k1}{\leftrightarrow}} ES \overset{kcat}{\rightarrow} E + P, \qquad (2\text{-}6)$$

where E is the enzyme, ES the enzyme-substrate complex, and P is the product. The constants k_1, k_{-1}, k_{cat} , determine the kinetics. The rate of each elementary step is determined by an energy barrier. Qualitative results do not change for the parameter ranges used (Fig. 7 e,f). Energy barriers have been varied between 0 k_BT and 4 k_BT except for the catalysis step for the phosphatase inactivating a kinase which was varied from $0k_BT$ to 8 k_BT . Note that changes in energy barriers alter the effective rate constants exponentially,

$$k \sim e^{-(E/k_BT)},$$

and so our parameter variations change the rates by factors of ~50 and ~3000 as the energy barriers are changed by 4 k_BT and 8 k_BT, respectively. The insensitivity of our results to such large variations comes from the fact that these parameter choices do not

limit the availability of the enzyme. Note that the effects of enzyme saturation[34] and distributive phosphorylation[48] have not been investigated.

Chapter 2.5.6 Computation of ensemble averaged quantities

All calculations presented are carried out at steady-state, and ensemble averaged quantities are reported. Ensemble averaging was performed by first allowing the system to approach steady state as determined by no time dependence in the mean square displacement,

$$\frac{d<[R(t)-R(0)]^2>}{dt}=0, \qquad (2\text{-}7)$$

and also ensuring that the simulation has advanced to where the time is much longer than the reaction time scales; sampling of each configuration is done thereon. Simulations are carried out for roughly 10^9 Monte Carlo steps, a time much longer than the time to approach steady-state and a time needed to acquire adequate statistics. These extended Monte Carlo trajectories ensure that there are no artifacts in our calculations due to insufficient statistical sampling. Analysis of the Monte Carlo trajectories shows that the approach to steady state is monotonic and that the results obtained are independent of initial conditions provided that the scaffolds are uniformly distributed within the simulation box. This indicates that the approach to steady state will also have the same properties as those reported for the steady state ensemble averaged quantities. Thus, dynamical properties of the signaling cascade can be inferred from the steady-state distributions that we calculate.

Computer code for the kinetic Monte Carlo simulations was written in ANSI C and compiled with the GNU C Compiler. Simulations were carried out with serially

clustered AMD Opteron 248 Processors. All data analysis was performed either with MATLAB or with code written in the PERL scripting language.

Chapter 2.5.7 Why ordinary differential equations (ODEs) were not used

The most commonly used method for modeling cell signaling dynamics is to use a set of ODEs and impose mass-action kinetic laws that govern the reaction dynamics subject to a prescribed network topology. It is therefore important to note that a model of scaffold-mediated signal transduction, based on ODEs and mass action kinetics, can not capture the physical effects of the scaffolding revealed by our studies. As discussed in the text, scaffold proteins impose a stoichiometric constraint on how many downstream targets with which an activated kinase can interact. As shown before in many contexts (e.g. [49, 50]), stoichiometric constraints introduce a length scale (L_{scaf} in our case). An ODE based model using the "well-stirred chemical reactor approximation"[51], on the other hand, is a model with infinite ranged interactions and no length scales. Scenarios exist (e.g. low phosphatases levels in our model) where the ODE model would predict that scaffolds enhance signal transduction when this is physically impossible because of the inhibitory effects of the spatial constraints imposed by the stoichiometric limitation.

Chapter 2.5.8 A possible model based on partial differential equations

We note that a different type of mean-field model involving a set of partial differential equations could conceivably capture the physical effects contained in the kinetic Monte Carlo simulations provided that the kinase concentration fields are

properly constrained in the model. This approach would involve the numerical solution of many coupled nonlinear PDEs. It may be interesting to investigate the qualitative behavior of a minimal model of this sort.

For an example, in a system with N scaffolds, and reactions characterized by kinetic parameters, k_+ (rate of phosphorylation), k_- (rate of dephosphorylation), k_{on} (rate of binding to the scaffold), k_{off} (disassociation rate for unbinding from the scaffold), and D (a protein diffusion coefficient), a set of reaction-diffusion equations can be constructed for the time evolution of the concentration of the i^{th} activated species (denoted in the superscript), $\rho_u^{*,i}(r,t)$, where the * indicates an activated kinase and the subscripts b and u identify a bound or unbound form of the molecule.

$$\begin{aligned}\frac{\partial \rho_u^{*,i}(r,t)}{\partial t} &= D\nabla^2 \rho_u^{*,i}(r,t) + k_+ \rho_u^{*,i-1}(r,t)\rho_u^i(r,t) - k_- \rho_u^{*,i}(r,t) \\ &\quad - \sum_{j=1}^{N} \delta(r-r_j)\begin{bmatrix} k_{on}\rho_u^{*,i}(r,t)\Theta(\rho^T - \rho_b^{*,i}(r_j,t) \\ -\rho_b^i(r_j,t)) - k_{off}\rho_b^{*,i}(r,t) \end{bmatrix}\end{aligned} \tag{2-8}$$

$$\frac{\partial \rho_b^{*,i}(r,t)}{\partial t} = \sum_{j=1}^{N} \delta(r-r_j)\begin{bmatrix} k_{on}\rho_u^{*,i}(r,t)\Theta(\rho^T - \rho_b^{*,i}(r_j,t) - \rho_b^i(r_j,t)) \\ -k_{off}\rho_b^{*,i}(r,t) \\ +k_+\rho_b^{*,i-1}(r,t)\rho_b^i(r,t) \\ -k_-\rho_b^{*,i}(r,t) \end{bmatrix} \tag{2-9}$$

where r_j is the location of the j^{th} scaffold, $\delta(r)$ is the Dirac delta function that confines each scaffold to a point in space, r, and the Heaviside step function, $\Theta(X)$ (defined as 0 for $X < 0$ and 1 for $X > 0$), imposes the constraint that the total concentration of a particular kinase $\rho^i{}_{b,u}(r,t) + \rho^{*,i}{}_{b,u}(r,t)$ can accumulate at each scaffold site at r_j only up to a threshold value, ρ^T.

In such a phenomenological model, the time evolution of kinase concentration fields by incorporating diffusion, activation by a downstream kinases, deactivation, binding and unbinding to the scaffold and inactivation and deactivation on the scaffold.

Chapter 2.6. Appendix

Chapter 2.6.1 Supplementary figures

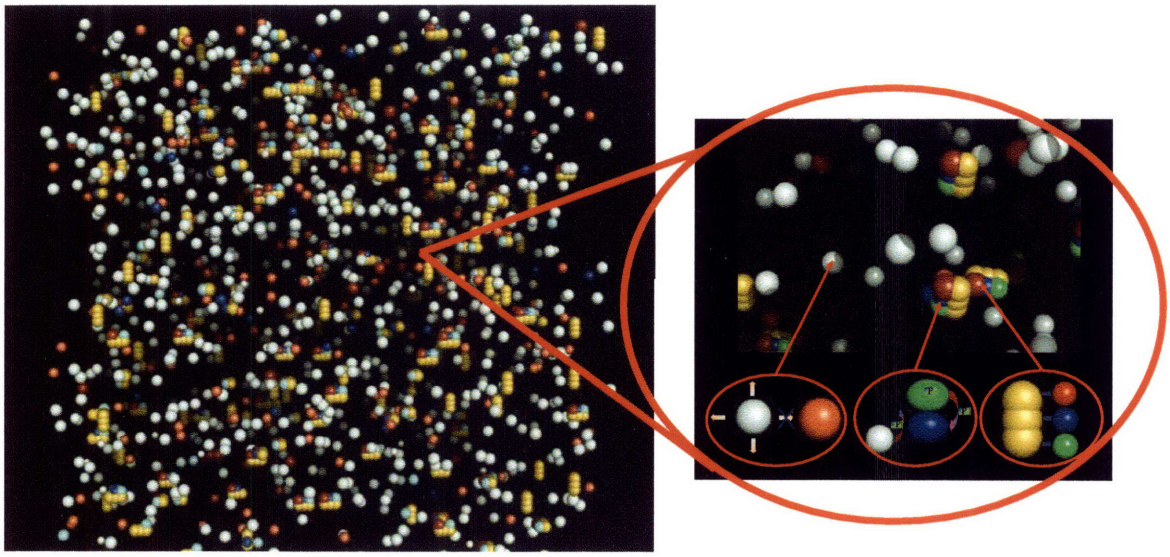


Figure 2-6. **A snapshot of the lattice Monte Carlo simulation**

A snapshot from the computer simulation with cartoons of the distinct types of events considered in our Kinetic Monte Carlo simulation. The stimulus, kinases, phosphatases, and scaffold proteins are represented as discrete objects on a lattice. The proteins can hop randomly to adjacent sites unless a target site is occupied by another protein. Each kinase can potentially be activated by an upstream kinase if the appropriate upstream kinase is activated and located at a proximal site. The rate of phosphorylation, determined by the activation barrier for the reaction, is varied. Phosphatases, upon encountering active kinases, can deactivate the kinase. Kinases can bind to sites on the scaffold protein with a binding energy, E_2, and

disassociate from its scaffold binding site with energy, E.

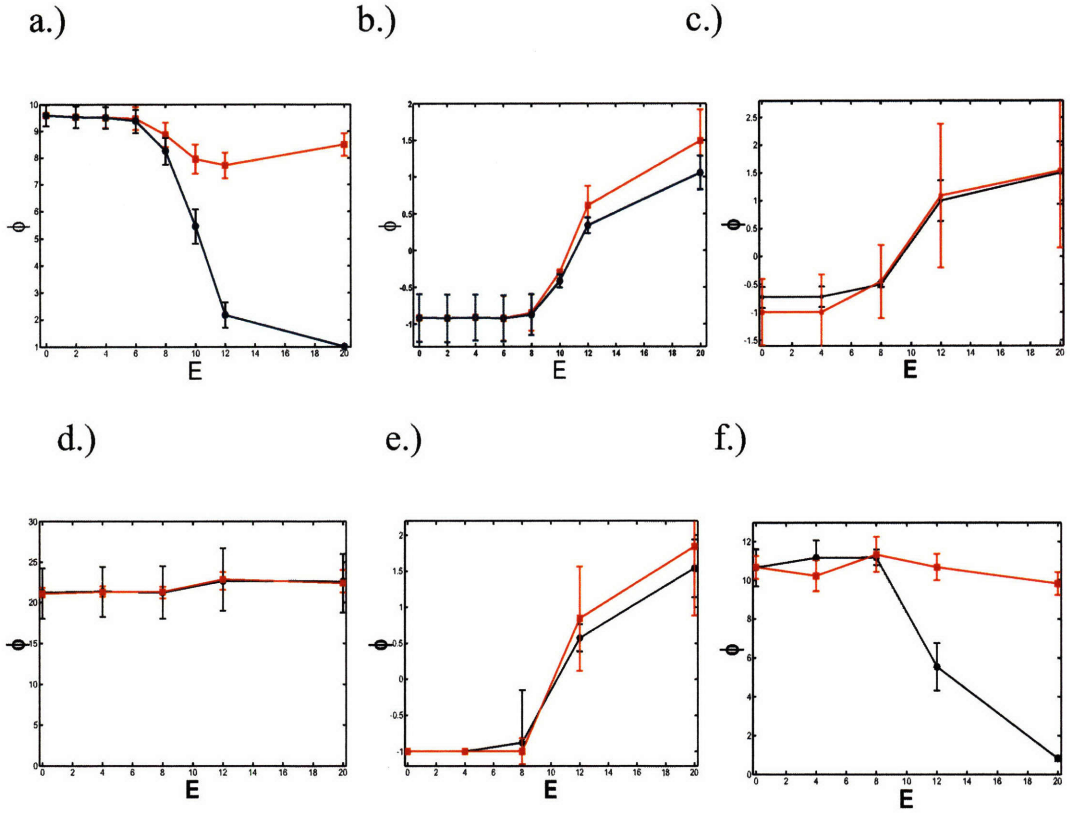


Figure 2-7. **Effects of kinase numbers and catalysis rates**

Plots of ($\phi \equiv \left\langle \frac{C^*}{A^*} - 1 \right\rangle$ for different values of kinases and kinetic parameters. Black lines represent the constrained case, red lines indicates the unconstrained case. (a) and (b): The ratio of A:B:C:scaffold is 1:1:10:1. (c) and (d) This concentration ratio is 1:10:10:1. (a) and (c) show cases of high phosphatase activity. (b) and (d) show cases of low phosphatase activity. $\sigma = 0.25$ and 1200 generic phosphatases are present in panels a-d. For concentration ratios of 1:1:10:1, the results are qualitatively the same as in Fig. 2. For concentration ratios of 1:10:10:1, results do not change qualitatively for high phosphatase activity. However , for low phosphatase activity (d), scaffolding the cascade has little effect on signaling. This is because kinases in free solution do almost all of the signaling in both the scaffolded and non scaffolded scenarios.
Qualitative results do not change upon variation of catalytic rates (e and f). (e) $E_3 = 4$ and $E_4 = 4$. (f) $E_3 = 4$ and $E_4 = 10$.

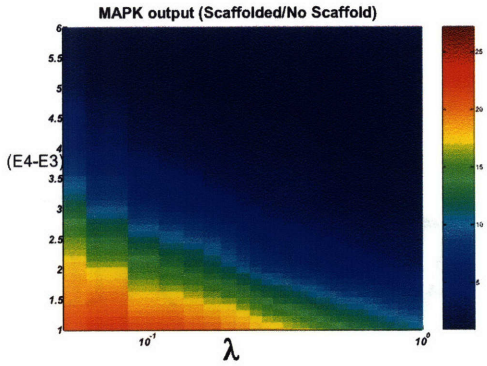


Figure 2-8. **How simulation results can depend on protein mobility**

On the abscissa, λ ($\lambda \equiv \frac{P_{Diffusion}}{P_{Reaction}}$), which sets a relative protein mobility is varied. The ordinate represents the difference in phosphatase deactivation energy and kinase phosphorylation energy barrier E_4-E_3, and $E_3 = 0$. We calculate the ratio of active kinases in the presence of equimolar amounts of scaffold protein with a strong binding affinity ($E_2 = 20$) to the number of active kinases in an unscaffolded system ($E_2 = 0$). For this particular calculation, a 1:1:1:1 ratio of kinases and scaffolds is used along with 800 generic phosphatases, and a strong stimulus is present ($\sigma = 0.25$). In this case, scaffolding the signaling system becomes more (less) important as phosphatase activity increases (decreases) and protein diffusion decreases (increases).

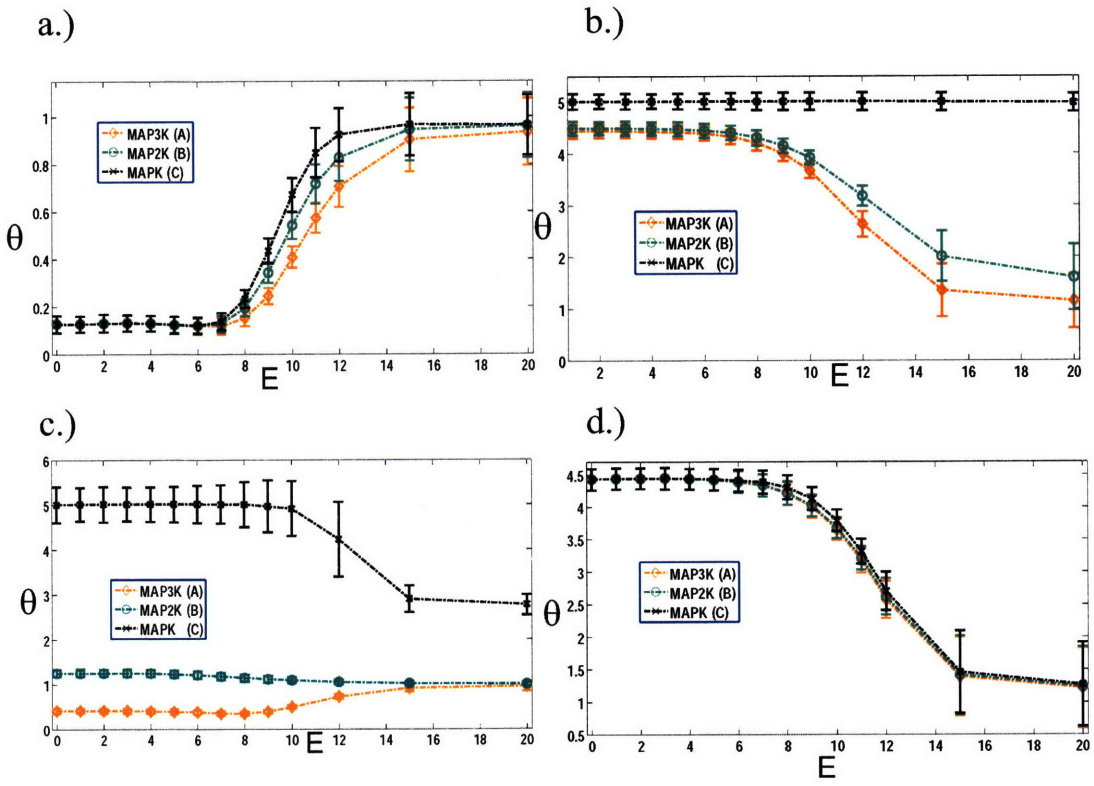


Figure 2-9. **Differential phosphatase kinase-phosphatase interactions can influence the signaling function of scaffold proteins**

Three situations are considered where the first (MAP3K (A), diamonds), second (MAP2K (B), circles), and third(MAPK (C) crosses) members of the cascade are protected from any phophatase mediated inactivation. In (a,b), the specified kinase, A (MAP3K), B (MAP2K), or C (MAPK) is protected from phosphatase-mediated deactivation only when bound to the scaffold. In (c,d), the specified kinase is protected from phosphatases regardless of whether or not it is attached to the scaffold. Kinases bound to the scaffold can not interact with their downstream substrates that are present in solution. Again high (a,c) and low (b,d) constitutive phosphatase levels are studied. Signal output, θ ($\theta \equiv \left\langle \frac{C^*}{[A]_0} \right\rangle$), is plotted again varying kinase-scaffold disassociation energies, E.

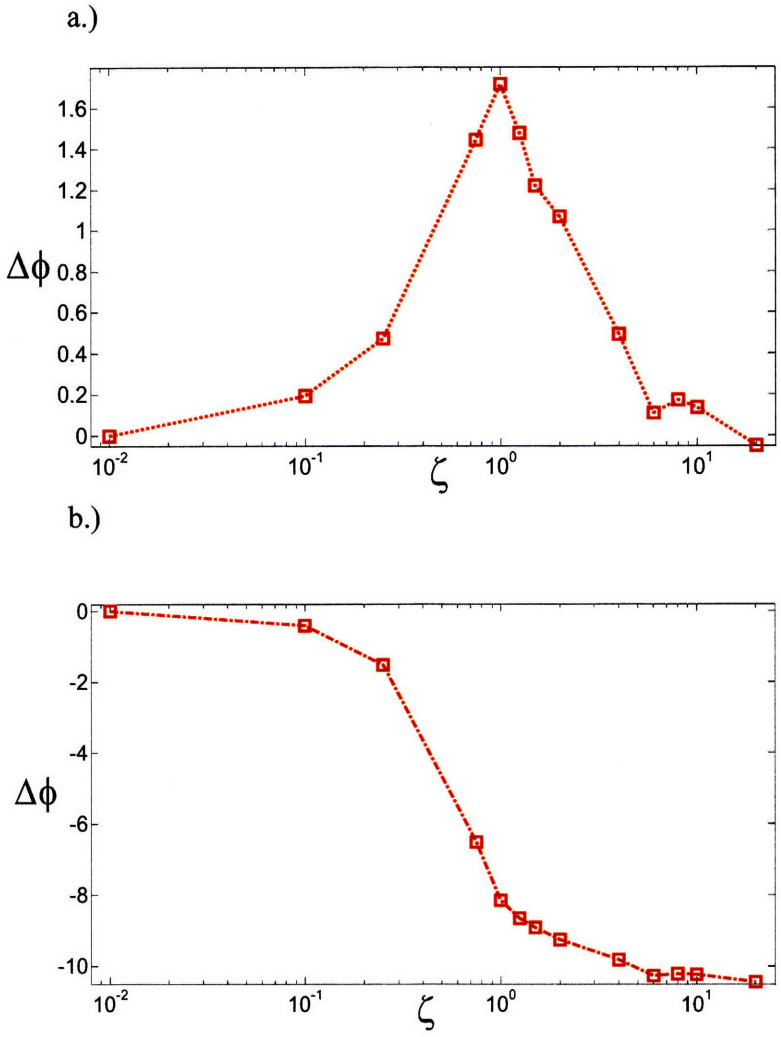


Figure 2-10. **Variations in scaffold concentration**

Shown below are representative results for how the difference in the amplification factor, $\Delta\varphi$ ($\phi \equiv \left\langle \frac{C^*}{A^*} - 1 \right\rangle$), for scaffolded (and tightly bound E = 20) and non scaffolded signaling depends upon relative scaffold concentration, ξ ($\xi \equiv \frac{[Scaffold]}{[A]_0}$). The ratio of kinase concentrations is 1:1:10, and $\sigma = 0.25$ ($\sigma \equiv \frac{[S^*]}{[A]_0}$). High (a) and low (b) constitutive phosphatase levels are considered. a.) As expected, amplification effects become prominent when a significant number of scaffolds are present and are most pronounced at $\zeta = 1$, the optimal value of scaffold concentration; however, amplification effects persist for scaffold concentrations

roughly 75% greater than the optimal value. Beyond this threshold value of scaffold expression, the prozone effect is dominant and amplification is abrogated. b.) In addition to the inhibitory prozone effect at high scaffold concentrations, under these conditions, scaffolds are inhibitory at all concentrations, and thus, increases in the scaffold concentration results in a monotonic decrease in signal output

Chapter 3 – Scaffold-mediated Signaling Dynamics: Regulation of Signal Duration and the Statistical Dynamics of Kinase Activation by Scaffold Proteins

Chapter 3.1 Introduction

In the context of signal transduction, cells must integrate signals derived from membrane proximal events and convert them into the appropriate cell decision. Within the complex networks that integrate these signals lies a highly conserved motif involving the sequential activation of multiple protein kinases. Signal propagation through these kinase cascades is often guided by a scaffolding protein that assembles protein kinases into a multi-protein complex. Signaling complexes maintained by scaffolds are intensely studied and have been shown to affect myriad cell decisions[8, 24-27, 52, 53]. Despite numerous advances in the understanding of the signaling function of scaffold proteins[30, 41, 42, 45, 54-57], many questions remain. For instance, although scaffolds are believed to have profound effects on the dynamics of signal propagation[8, 31, 54, 55], the mechanisms that underlie how scaffolds regulate signaling dynamics are not well understood.

One key factor in specifying a cellular decision is the duration of a signal (i.e. the time over which a kinase remains active)[58, 59]. Differences in signal duration have been implicated as the basis of differential decisions in myriad cell processes. For example, it has been suggested that decisions on growth factor induced cell proliferation, positive and negative selection of T cells, apoptotic programs, cell cycle progression, among many others, are regulated by the duration of signaling[60-65]. Therefore, the issue of how a signal output, such as the activity of extracellular regulatory kinase (ERK) in a MAPK pathway, is distributed over time, is of considerable interest.

There are many ways in which the duration of the output of a kinase cascade can be controlled. Regulation of signaling dynamics can arise from processes upstream of the

cascade[66]. For example, degradation of upstream signaling components such as the surface receptors[67] and differential kinetics of GTPase regulators[68, 69] can be essential in regulating MAPK signaling dynamics[66]. It has been also been shown that differential modes of feedback regulation that are manifested under different conditions within the same cascade can regulate signal duration[70]. Scaffold proteins have also been implicated as key determinants in the regulation of signal duration[54, 55, 71].

Because the many factors that control scaffold mediated signaling are difficult to systematically control in a laboratory setting, a precise understanding of how scaffold proteins affect the dynamics of signal transduction has proven elusive. Computational models have been useful in understanding some of the many complex ways in which scaffolds influence signal transduction[12, 29, 31, 72]. However, it is currently impossible to model theoretically all aspects of any biological signaling process—computational models ultimately require that many gross simplifications be made. Our aim is, therefore, not to attempt to simulate every detail of a specific biochemical pathway but rather investigate the consequences that emerge from a simple scenario of scaffold mediated signaling whereby a model cascade assembles onto a scaffold. In modeling this scenario in itself, we hope to learn more about the functional and mechanistic consequences that these specific physical constraints, imposed by assembling components of a biochemical cascade onto a scaffold, confer to signaling pathways. In parsing these effects from the myriad others that are undoubtedly important, our hope is that our results can serve as a framework for understanding the extent to which these effects are important in specific biological contexts such as the Mitogen Activated Protein Kinase (MAPK) pathway.

One theoretical analysis of scaffold mediated cell signaling revealed the presence of non mononotic behavior in signal output as a function of scaffold concentration[29]. If scaffolds are required for signaling, then too few scaffolds will be detrimental to signaling. On the other hand, if scaffolds are present in excess, signaling complexes become incompletely assembled and the signal output is attenuated. As a consequence of this "prozone" effect, scaffolds were shown to also differentially affect the kinetics of signaling.

The observation that scaffolds can differentially affect signaling dynamics leads to many questions. How do scaffold proteins control the time scales involved in signal propagation? An important metric of cell signaling is the time it takes for a downstream kinase to become active[73, 74]. As signal transduction is stochastic in nature, the more precise question is: what is the distribution of times characterizing the activation of a downstream kinase? How do scaffolds affect this distribution, and what might be the biological consequences of changes in this distribution as a result of signaling on a scaffold? We compute first passage time distributions[75] using a stochastic computer simulation method to investigate these questions.

Specifically, we use a kinetic Monte Carlo algorithm. We have previously used such methods to study a different question concerning the regulation of signal amplitude by scaffold proteins[12]. It is also possible that a differential equation model that considers mean-field kinetics could be used to study the first passage time distribution [75]. However, such an approach would require the imposition of absorbing boundary conditions that can make the numerical analysis difficult.

Our simulation results suggest that, depending on physiological conditions, scaffold proteins can allow kinase cascades to operate in different dynamical regimes that allow for large increases and decreases in the speed and characteristic time scale of signal propagation. Furthermore, and perhaps more importantly, scaffolds are shown to influence the statistical properties of the times at which kinases are activated in complex ways. Scaffolding protein kinases cascades can allow for broadly distributed waiting times of kinase activation, whereas in the absence of a scaffold, the time it takes for a kinase to be activated is effectively characterized by a single time scale. These stochastic characteristics of scaffold-mediated kinase cascades are, to our knowledge, elucidated for the first time and may have diverse biological consequences that pertain to how signal duration is regulated. It is also our hope that our results provide a framework for achieving a deeper qualitative understanding of how scaffolding proteins can regulate the dynamics of cell signaling and the statistical properties of signal transduction.

Chapter 3.2. Results

Chapter 3.2.1 Model of a protein kinase cascade

For our study, we considered a model three tiered protein kinase cascade such as the MAPK pathway[19]. Since our aim is to study the effects of spatially localizing protein kinases on signaling dynamics, we considered a minimal description of a model kinase cascade. Many factors that are undoubtedly important in regulating signaling dynamics were not considered. These factors include feedback regulation within the cascade, allosteric and or catalytic functions provided by the scaffold, and the effects of multiple phosphorylations of each kinase[23, 30, 48, 66, 67, 70].

In our model, signal propagation occurs in a three step hierarchical fashion: an initial stimulus (S) activates a MAP3K (A) that in turn, activates a MAP2K (B), that subsequently can activate its MAPK (C) substrate—phosphatases can deactivate each activated species and this deactivation occurs regardless of whether or not the active kinase is bound to a scaffold. A schematic is presented in Fig. 1a that illustrates the basic processes that are allowed in our model. A steady-state ensemble is considered. That is, simulations are allowed to first reach a dynamic steady-state and once this state is reached, dynamics are studied. We do not consider dynamics from the starting time that requires propagation through a hierarchical cascade.

Recent work has studied the statistical dynamics of kinase activation that result from the hierarchical organization of a kinase cascade; in that study, it was shown that the hierarchical structure of the cascade gives rise to broad waiting time distributions of cascade activation. In the regime that we study here, these effects are absent since activation of the cascade requires that an inactive C protein encounter an active B protein; our motivation is thus to investigate how the dynamics of kinase activation can be affected by assembling components of the cascade onto a scaffolding protein that localizes single complexes. Therefore, we do not emphasize how the hierarchical structure of a signaling cascade effects signal propagation and instead focus on how assembly of the cascade onto a scaffold affects signaling dynamics. We also underscore the notion that in our approach, many undoubtedly important effects such as the hierarchical structure of protein kinase cascades, the influence of feedback loops, differential enzymatic mechanisms and allosteric control by scaffolds are neglected. Again, by excising these effects, we restrict our attention to a hypothetical scenario that

aims only to investigate the consequences of assembling components of a cascade onto a scaffold protein.

The key quantities computed and parameters used are discussed below, in Table 1, and in Fig. 1b. Additional details are provided in the methods section.

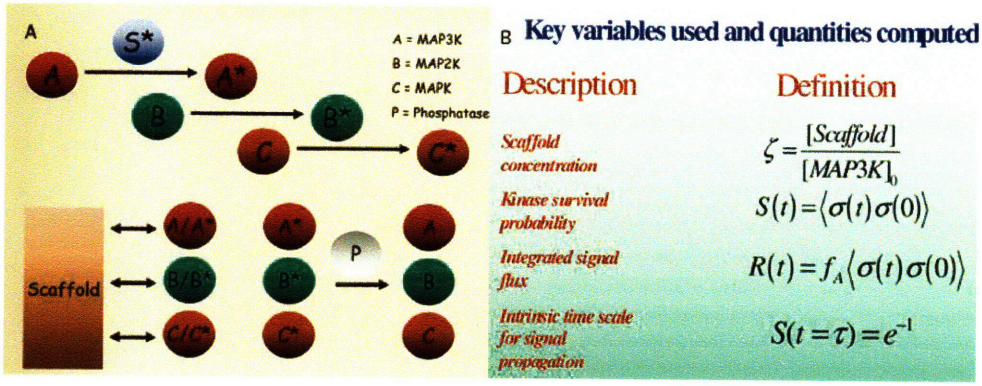


Figure 3-1. **A model to study dynamical properties of a scaffold mediated signaling cascade**

a.) Schematic of the events considered in the scaffold mediated signaling cascade. Each kinase, if activated, can activate its downstream substrate when the two proteins are in close proximity. Kinases can bind and unbind to the scaffold and phosphatases can, upon encountering an active kinase, deactivate it. Activation potentially occurs both in solution and on a scaffold. Each forward and backward reaction is modeled as a elementary reactive collision with an energy barrier, E. Energy barriers, E, were taken to be zero so that all kinetics are diffusion limited b.) Key variables and the main quantities computed are shown. ζ is the dimensionless scaffold concentration. The concentration of scaffold proteins $[Scaffold]$ is scaled to the density of the first kinase in the cascade $[MAP3K]_0$ ($\zeta \equiv \frac{[Scaffold]}{[MAP3K]_0}$). The survival probability $S(t) \equiv \langle \sigma(t)\sigma(0) \rangle$, where $\sigma(t)$ is zero is a kinase has become active and one otherwise (brackets denote an ensemble average); $S(t)$ is the probability that the final kinase has been activated given that it was inactive at $t = 0$.
$R(t) \equiv f_A S(t) = f_A \langle \sigma(t)\sigma(0) \rangle$, where f_A is the fraction of active kinases at steady state; $R(t)$ is the integrated reactive flux of

kinase activation. The characteristic time scale of signal propagation τ is defined by the relation $S(t=\tau)=e^{-1}$.

Chapter 3.2.2 The concentration of scaffold proteins sets time scales for signal propagation through a kinase cascade

To set the context, consider the consequences of signaling in two limiting cases in our model. When the binding affinity of the kinases to the scaffold, E, is low (defined here to be close to the thermal energy, $E \sim k_b T$) and kinases disassociate rapidly from the scaffold, few proteins on average are bound to a scaffold. Therefore, signaling dynamics corresponds to that of a kinase cascade in solution. For a very strong affinity, $E >> k_b T$, all available binding sites to scaffold proteins are occupied by kinases (on average). In this case, signaling dynamics are controlled by the time required for initial stimuli to encounter and interact with each fully assembled complex.

Therefore, we consider cases in which kinases can disassociate from their scaffolds and exchange with unbound kinases on time scales pertinent to cell signaling processes. Such time scales correspond to disassociation constants (K_d) on the order of 1-10 μM and off rates, $k_{off} \sim 1 s^{-1}$. Such K_d values correspond to free energies of binding of roughly 7-9 kcal/mol, an energy scale typical of protein-protein interactions in kinase cascades[76]. We have used $12 k_b T$ as the binding energy in our simulations which corresponds to $\sim 7.2\, kcal/mol$. We also discuss the robustness of our results with respect to changes in this value. Scaffold concentration has been identified as a key variable that can regulate the efficiency of signal propagation through a kinase cascade[24, 25, 29]. For the set of parameters used in the simulations (Table 1.), signal output (defined as the

average steady state value of the final kinase in the cascade) has a non-monotonic (biphasic) dependence on the relative concentration of scaffolds ζ ($\left(\zeta \equiv \frac{[Scaffold]}{[MAP3K]_0}\right)$, where $[Scaffold]$ is the concentration of the scaffold and $[MAP3K]_0$ is the concentration of the first kinase in the cascade) and peaks at an optimal value of $\zeta = 1$ [12, 29].

To quantify signaling dynamics, we consider a survival probability $S(t)$ (methods) that, as mentioned, can be viewed as a type of autocorrelation function.

$$S(t) \sim \langle \sigma(t)\sigma(0) \rangle, \qquad (3\text{-}1)$$

where $\sigma(t)$ equals 0 or 1 depending upon the activity of the final kinase within the cascade (methods) and the brackets indicate an average over all kinases in the simulation averaged over many simulations. This quantity gives the probability that the final kinase in the cascade remains inactive at time t given that it was inactive at time $t = 0$. Therefore, signaling dynamics can be monitored by observing the decay of this function with time.

In Fig. 3-2a, $S(t)$ is computed for different values of the relative scaffold concentration, ζ. The intrinsic time of signal propagation, τ, is the value at which $S(t)$ decays to e^{-1} of its original value($S(t = \tau) = e^{-1}$). Upon increasing scaffold concentration, τ increases. At very high scaffold expression levels, signals propagate so slowly that cell signaling is not observed on experimentally measurable time scales which we take to be in our simulations >> 10^6 Monte Carlo (MC) steps; 1 MC step ~ 1μs assuming a lattice spacing of 10 nm and a diffusion coefficient of $10 \mu m^2/s$ [33]. The increase in τ spans several orders of magnitude as is observed in Fig. 3-2b. Distinct

stages are also observed in the behavior of τ, and are separated by an inflection point occurring shortly past the optimal value of scaffold concentration ($\zeta \sim 1$). This phenomenon suggests that different physical processes are determining the signaling dynamics at different ranges of scaffold concentration.

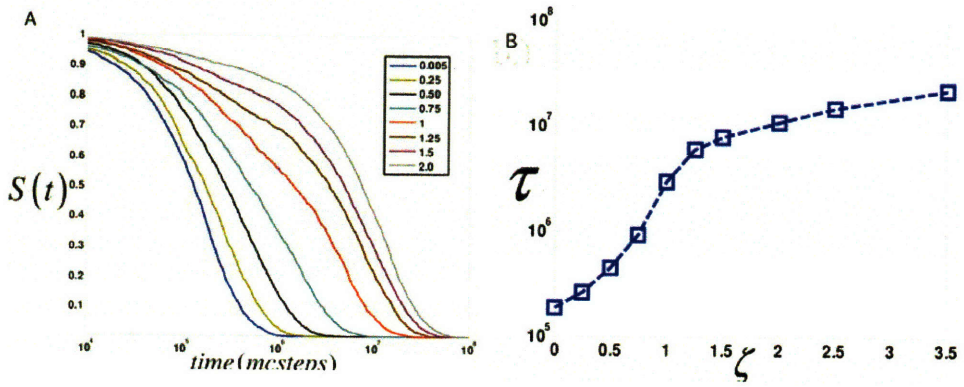


Figure 3-2. **The concentration of scaffold proteins sets time scales for signal propagation.**

a.) $S(t)$ as a function of time for different values of relative scaffold concentration, ζ ($\zeta \equiv \frac{[Scaffold]}{[MAP3K]_0}$). ζ ranges from 0.005 to 2 times the optimal value. b.) The characteristic time scale τ ($S(t=\tau) = e^{-1}$) is extracted from the curves in (Fig. 2a.), and its variation with ζ is shown. Two regimes are observed and are separated by an inflection point. c.) Integrated signal flux $R(t)$ for different values of ζ.

These results also suggest that the concentration of scaffold proteins can in principle set an intrinsic time scale that determines the speed of signal propagation. Such an intrinsic time scale arises solely from changes in the concentration of scaffold proteins. This time scale can span several orders of magnitude for biologically relevant affinities and diffusion coefficients and increases monotonically with increasing scaffold concentration.

Note that these calculations consider only the speed of signaling and do not necessarily imply that signaling is more efficient when τ is small. To observe the total amount of integrated signal flux, the survival probability is conditioned with the probability that a kinase in the pool of signaling molecules is active in the steady state. We compute $R(t)$ defined as $S(t)$ multiplied by the average number of (the final downstream) kinases active at steady state,

$$R(t) \equiv f_A \langle \sigma(t)\sigma(0) \rangle, \qquad (3\text{-}2)$$

where f_A is the fraction of active kinases at steady state). The time derivative, $-\frac{d}{dt}R(t)$, can be thought of as a flux of activated kinases being produced. In Fig. 3-2c, $R(t)$ is plotted as a function of time. For low concentrations of scaffolds, the small amount of signal, albeit quickly propagating, is rapidly quenched. As scaffold concentration increases, both the amplitude and duration of the signal increase up to an optimal value. Past the optimal value, higher scaffold concentrations result in signals with small amplitude but the duration of signaling is extended. The behavior of the integrated reactive flux is a direct consequence of the existence of an optimal scaffold concentration and "bell shaped" titration curve since the area under these curves is proportional to the average signal output[12, 29].

Chapter 3.2.3 Scaffold proteins influence the duration of signaling by controlling how kinase activation is distributed over time

Fig. 3-2. emphasizes how the characteristic time for signal propagation is influenced by changes in the relative scaffold concentration. It also appears that the qualitative features of $S(t)$ change as scaffold concentration is varied. The decay of some distributions appears highly concentrated at a particular time while the decay of other distributions appears more broadly distributed.

To further investigate this observation, we plotted the survival probability as a function of the dimensionless time, t/τ. If the decay of $S(t)$ is purely exponential, then $S(t/\tau)$ will have the form $e^{-t/\tau}$. Fig. 3 shows $S(t/\tau)$ for different values of scaffold concentration and a decaying exponential function is given as a reference. One notices that $S(t/\tau)$ is exponential at negligible scaffold concentrations. As scaffold concentration increases, the behavior of $S(t/\tau)$ deviates from a single exponential decay. Near $\zeta = 1$, $S(t/\tau)$ shows maximal deviation from purely exponential kinetics. As scaffold expression increases past this point, the shape of $S(t/\tau)$ reverts back to an exponential form.

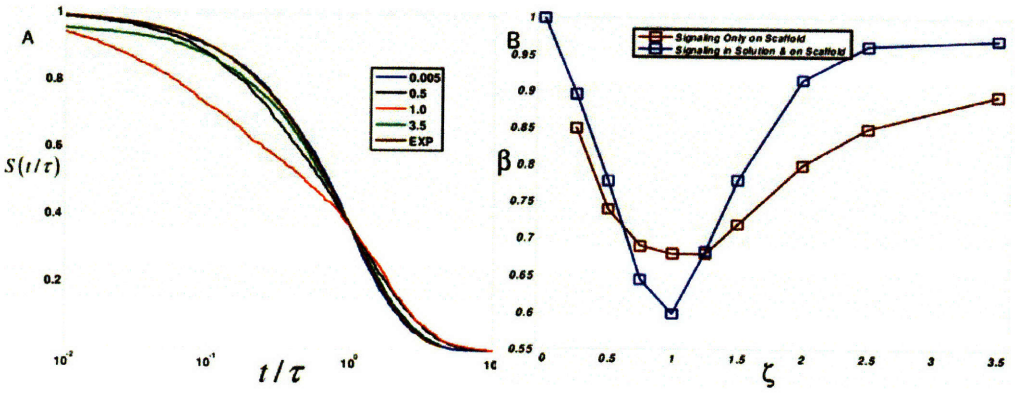


Figure 3-3. **Scaffold proteins allow for signals to be distributed over many time scales.**

Variation of the survival probability with time scaled to characteristic time scales, τ. a.) $S(t/\tau)$ (on a semi-logarithmic scale) for different values of ζ. Values of ζ are given in the legend. Large deviations of exponential decay are observed near the optimal value of ζ (ζ =1, red). b.) Survival probabilities were fit to a stretched exponential function ($S(t/\tau) \sim e^{(t/\tau)^{\beta}}$). Values of stretching exponent β as a function of scaffold density ζ are shown. Two cases are considered: (1) kinases can be activated only while bound to a scaffold (red) and (2) kinases can be activated while in solution and bound to a scaffold (blue). β deviates most from a purely an exponential (β = 1) at the optimal value of scaffold density (ζ =1) in both cases.

A deviation from exponential behavior can be quantified by considering a stretched exponential function,

$$e^{-(t/\tau)^{\beta}}, \qquad (3\text{-}3)$$

and fitting $S(t/\tau)$ to this form for different values of ζ. One desirable feature of the stretched exponential function is the minimal number of parameters, τ and β, that are involved in the least-squares fit; also, the values of these parameters can be physically interpreted. τ gives the characteristic time for one overall timescale of signal propagation, and β is a measure of how much the function, $S(t)$, deviates from a single exponential and thus how broadly distributed are the signaling dynamics. Fig. 3b shows how β depends on scaffold concentration. For these simulations, $\beta \sim 1$, for small and large values of scaffold concentrations indicating exponential behavior. For intermediate values, β peaks at a minimum of $\beta \sim 0.6$, a significant deviation from purely exponential behavior.

In the limits of small and large scaffold concentrations, the presence of a single exponential decay, $\beta \sim 1$, indicates that signal propagation, or the relaxation of $S(t/\tau)$,

occurs at one characteristic time scale. In the intermediate regime, β shows significant deviations from one, thus allowing for a broadly distributed signal. When β is significantly less than one, signals can steadily propagate over several decades. In this regime, the waiting time distribution $f(t)$,

$$f(t) \equiv -\frac{dS(t)}{dt} = \frac{\beta}{t}\left(\frac{t}{\tau}\right)^{\beta} e^{-(t/\tau)^{\beta}}, \qquad (3\text{-}4)$$

has a large tail and the activation of kinases is slowly maintained over many time scales.

Chapter 3.2.4 A multi state kinetic mechanism illustrates the competition between the many time scales involved in scaffold mediated cell signaling

Why do we observe exponential and non-exponential behavior under different conditions? Signal transduction in our model occurs on a time scale that is much slower than the microscopic time scales associated with diffusion, binding/unbinding, and enzyme catalysis. We might therefore expect that some coarse-graining exists whereby events at these fast, "microscopic" time scales interact with other relevant biophysical parameters (e.g. scaffold concentration) to give rise to emergent properties that evolve on slower times scales. These processes are a manifestation of the collective dynamics of the many processes that occur on faster time scales. Understanding the factors that govern these emergent time scales would then provide insight into the origin of the different temporal characteristics that are revealed by our simulations.

In order for a signal to propagate (i.e. for the last kinase in the cascade to become active), a hierarchical sequence of phosphorylation reactions among kinases must occur

that leads to the final kinase in the cascade being activated by its upstream kinases. The activation process may occur either in solution or on a scaffold. Also, in the course of signaling, kinases can exchange from a scaffold. Some kinases are bound to a scaffold that contains an incomplete assembly of the necessary signaling molecules, and are not signaling competent. Ultimately, an inactive kinase can exist in one of three states: in solution, bound to a complete complex, or bound to an incomplete complex. Fig. 3-4a contains a diagram of such a minimal picture and arrows denote transitions between the four states.

This minimalist description clarifies the behavior in Figs. 3-3a and b. For low scaffold concentrations ($\zeta << 1$), kinases predominately exist in solution and signal transduction is dominated by the time it takes for an upstream kinase to encounter its downstream enzyme. Since a steady-state ensemble is used, the rate limiting step for signal propagation is the diffusion limited collision between an active B* molecule with an inactive C molecule. For high scaffold concentrations ($\zeta >> 1$), kinases predominately exist in incomplete signaling complexes and signal transduction is limited by a time scale that characterizes the turnover of a signaling incompetent complex to one that is able to signal. For intermediate concentrations, inactive kinases can exist in each of three states and transitions between these states also occur. Thus, the source of the nonexponential relaxation (i.e. $\beta < 1$) arises from the mixing of many time scales that are relevant for intermediate scaffold concentrations. Fig. 3-4b illustrates this minimal picture of the kinetics of signal propagation derived from these physical considerations.

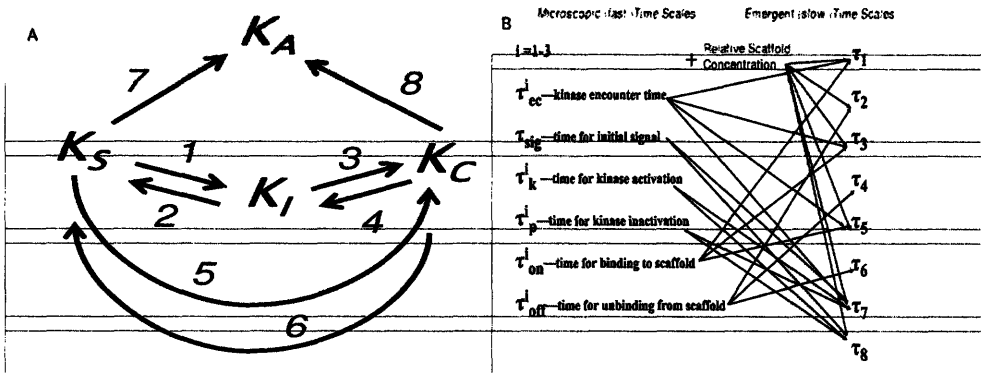


Figure 3-4. **Dynamics can be characterized by a multi-state kinetic mechanism.**

Important time scales in scaffold mediated signaling. a.) graph of multi-state kinetic model whose dynamics are governed by 8 transitions. Each kinase can transition between four states denoted with four subscripts: in solution (S), bound to a signaling competent complex (C), bound to a signaling incompetent complex (I), and activated (A). b.) Diagram depicting how the various processes occurring at fast time scales couple with scaffold concentration ζ to give rise to collective behavior occurring at slower time scales. Superscript i indicates kinases 1-3 in the cascade. Diffusion (τ^i_{ec}, τ_{sig}), catalysis ($\tau^i_k, \tau^i_p, \tau_{sig}$), and binding/unbinding from scaffolds ($\tau^i_{on}, \tau^i_{off}$) coupled to scaffold concentration to determine transitions among the four states involving 8 time scales ($\tau_i \in \{\tau_1, ..., \tau_8\}$).

Also note that the sensitivity of our results to changes in model parameters can be understood from this simple picture of scaffold mediated signaling dynamics. For instance, changes in kinase and scaffold concentrations result in changes in the relative amount of kinases existing in the three states in ways that have been previously characterized[12, 29]. Changes to other parameters such as the rates of activation and deactivation and the concentration of phosphatases alter the rates of transitions between these different states. For instance, if phosphatase concentrations are very large, then activation in solution is very slow and occurs predominantly on a scaffold. Also, slower

rates of activation (and larger rates of deactivation) result in a larger portion of signaling originating from kinases that are bound to scaffolds. In general, when the activation of kinases originates more (less) predominantly from a particular state in the minimal model, β increases (decreases). When multiple pathways to kinase activation contribute with comparable time scales, β is small, and signaling is broadly distributed over many time scales. We have performed many simulations with varying parameters to test the robustness and parameter sensitivity of our findings and find that that the qualitative behavior of our results follow this simple, qualitative, physical picture.

Additional insight can be gleaned from consideration of the power spectrum of $S(t)$. The power spectrum,

$$P(\omega) \equiv \left| C(\omega) \right|^2 \tag{3-5}$$

where,

$$C(\omega) = \int e^{i\omega t} S(t)\, dt \tag{3-6}$$

, computed in the frequency domain, resolves the time scale dependence of kinase activation. This approach has proven useful in studying the dynamics of complex biochemical networks in many contexts[77-79]. We first note that $S(t)$ obtained from the simulations fits well to the functional form $e^{-(t/\tau)^{\beta}}$ (χ^2 values small). Thus, we use the parameters β and τ that were extracted from the fits at low ($\zeta = 0.001$), optimal ($\zeta = 1.0$), and high ($\zeta = 3.5$) scaffold concentrations to compute $P(\omega)$ for these three cases.

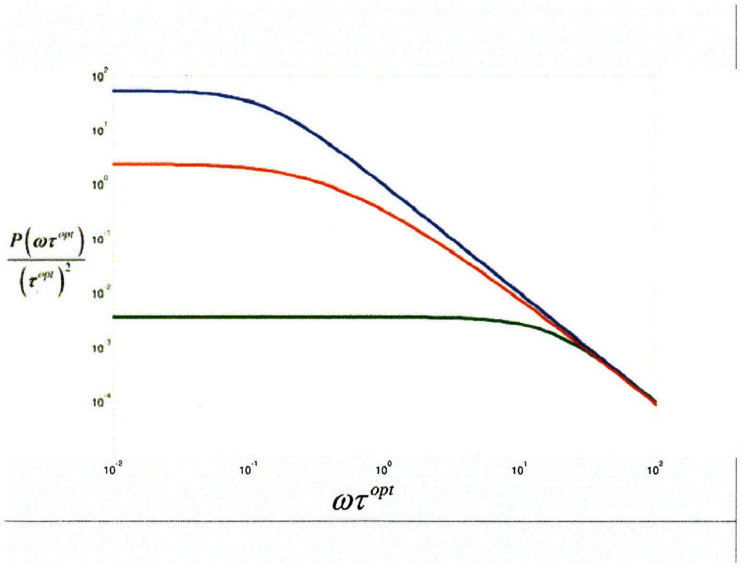


Figure 3-5. **Power spectra for kinase activation in the case of high, low, and optimal scaffold concentrations**

Plots of $P(\omega)=|C(\omega)|^2$; $C(\omega)=\int e^{i\omega t}S(t)dt$ where $S(t)=e^{-(t/\tau)^\beta}$, are considered. Three cases are considered: low concentration ($\zeta=0.005, \tau=1.9\times10^5\, mcsteps, \beta=1.03$), high concentration ($\zeta=3.5, \tau=2.2\times10^7\, mcsteps, \beta=0.97$), and optimal concentration ($\zeta=1.0, \tau=3.0\times10^6\, mcsteps, \beta=0.60$). On the x axis, frequency is reported in units that are scaled to the characteristic time for the $\zeta=1.0$ case, $\tau^{opt}=3.0\times10^6\, mcsteps$. The y axis contains values of $P(\omega\tau^{opt})/(\tau^{opt})^2$.

In Fig. 3-5., $(\tau^{opt})^{-2}\,P(\omega\tau^{opt})$ is plotted versus $\omega\tau^{opt}$ where the time τ^{opt} is the characteristic time scale τ for relaxation at the optimal $\zeta=1$ scaffold concentration. That is, time is rescaled to units of τ^{opt}. For each curve, at low $\omega\tau^{opt}<<1$ frequencies $P(\omega\tau^{opt})$ is constant ($P(\omega\tau\to0)\to\tau^2$) signifying that kinase activation has decorrelated. At high $\omega\tau^{opt}>>1$ frequencies, kinase activation is correlated and a power law decay is observed for each curve $P(\omega\tau^{opt})\sim\omega^{-2}$.. As a reference, note that for an

exponential decay, $S(t) = e^{-t/\tau}$, the transition between these two regimes occurs at $\omega\tau \sim 1$ and is determined by the Lorentzian:

$$P(\omega\tau) = \frac{\tau^2}{1+(\omega\tau)^2}. \qquad (3\text{-}6)$$

In Fig. 3-5, for high ($\zeta = 3.5$, blue) and low ($\zeta = 0.001$, green) scaffold concentrations power spectra closely resemble the Lorentzian with the transition to $P(\omega\tau^{opt}) \sim \omega^{-2}$ behavior occurring at different frequencies. At low $\zeta = 0.001$ concentrations, the inverse time scale or corner frequency at which kinase activation decorrelates is determined by the diffusion limited rates of activation and deactivation of the final kinase C*. The corner frequency can be estimated from

$$S(t) \sim e^{-(t/\tau_c)} \qquad (3\text{-}7)$$

, where,

$$(\tau_c)^{-1} = (k_+ + k_-) = O[D N_{tot} a]. \qquad (3\text{-}8)$$

k_+ and k_- are diffusion limited rates of activation and deactivation and are given by a diffusion limited encounter rate that is on the order of $D N_{tot} a$ where D is the diffusion constant used in the simulation, N_{tot} is the number of proteins, and a is the size of a protein taken to be the size of a lattice site. Substitution of the numbers used in the simulation (Table 1 and methods) achieves a value for the relaxation time that is commensurate with the relaxation time for $\zeta = 0.001$ in Fig. 3-2; i.e. $\tau_c \sim 10^5 mcsteps$. At high $\zeta = 3.5$ concentrations, the corner frequency is determined by rates of formation and disassociation of an intact signaling complex. Furthermore, because of these many

process that comprise the relaxation rate in this case, a numerical estimate of the corner frequency is difficult. In the case of the optimal ($\zeta = 1.0$, red) concentration, the transition from constant to $P\left(\omega\tau^{opt}\right) \sim \omega^{-2}$ behavior occurs smoothly over many decades from $\omega\tau^{opt} \sim 0.1$ to $\omega\tau^{opt} \sim 10.0$.

The plot in Fig. 3-5 also resolves different frequency dependent processes occurring in signal transduction. At high frequencies or short times, $\omega\tau^{opt} > 10.0$, kinase activation is limited by the diffusive motion of the kinases in the cascade. At intermediate frequencies, $0.1 < \omega\tau^{opt} < 10.0$, activation is dominated by transitions between kinases assembled in competent, incompetent, and solution based kinases. For low frequencies $\omega\tau^{opt} < 0.1$ or long times, kinase activation decorrelates for each scaffold concentration.

Chapter 3.2.5 Relationship between signal duration and computed first passage time statistics

To illustrate how computed values of $S(t)$ and the distribution of waiting times for kinase activation relate to conventional means of defining signal duration, we consider a differential equation for the time evolution of the activated form of the final kinase within the cascade. In this picture, species become activated at rates derived from the functional form that was fitted to the survival probabilities that were computed from the simulations. The waiting time or first-passage time distribution $f(t)$ is used as a forward rate and the activated final kinase then can be degraded with a kinetics of

degradation characterized by a rate constant, k_ϕ. A kinetic equation describing this process is written as:

$$\frac{dx}{dt} = \left\{ \frac{\beta(t/\tau)^\beta}{t} e^{-(t/\tau)^\beta} \right\} - k_\phi x . \qquad (3\text{-}8)$$

x is the number of active species, τ is the time constant of signal propagation and β is the stretching parameter that quantifies deviations away from single exponential behavior. In this picture, $x(t)$ represents the average response to a stimulus $f(t)$ that is distributed temporally according to $\frac{\beta(t/\tau)^\beta}{t} e^{-(t/\tau)^\beta}$ and subject to a first order decay with characteristic time $1/k_\phi$.

The equation for $x(t)$ can be solved and using the initial condition, $x(0) = 0$:

$$x(t) = \int_0^t e^{-k_\phi (t-t') - (t'/\tau)^\beta} \beta (t')^{-1} (t'/\tau)^\beta \, dt' . \qquad (3\text{-}9)$$

$x(t)$ was integrated numerically and is shown for different values of β in Fig. 3-6a. As seen in Fig. 3-6a, decreasing values of β result in the trajectories having longer tails and thus an extended duration of signaling. Also, smaller values of β result in the signal having a larger peak. This property directly follows from the decay of $S(t)$ that was shown in Fig. 3-3a for different values of β. At early times, $S(t)$ decays more quickly when β is smaller; as a consequence, more kinases are activated at these times, thus resulting in a larger peak.

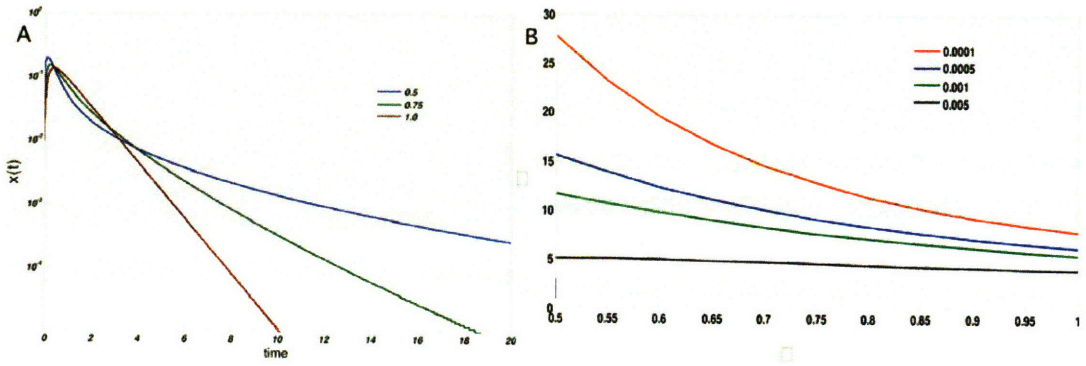


Figure 3-6. **Relation between first passage time statistics and signal duration**

a.) Trajectories of $x(t)$ on a semi-log plot. The abscissa represents time scaled by the characteristics time τ . Trajectories for three values of β are shown: $\beta = 0.6$ (blue), $\beta = 0.8$ (green), and $\beta = 1.0$ (red). $k_\phi = 5$ for each curve. Smaller values of β result in larger values of $x(t)$ at longer times. b.) Values of signal duration as a function of β for different choices of threshold, T (defined in text); values of T are provided in the legend. For small values of T , $\beta < 1$ (i.e. the presence of scaffolds) markedly increases signal duration.

This concept of signal duration can be made more precise by considering a threshold amount of signal, T , that is required for the pathway to be considered active. With a chosen value of T , the signal duration, υ , is defined as the time it takes for the signal to decay to some threshold value, T . That is, the equation,

$$T = \int_0^{\upsilon} e^{-k_\phi(\upsilon - t') - (t'/\tau)^\beta} \beta (t')^{-1} (t'/\tau)^\beta \, dt', \qquad (3\text{-}10)$$

is satisfied. Fig 3-6b. shows the signal duration, υ , as a function β for values of β ranging from 0.5 to 1 for different values of T . For smaller values of T , $\beta < 1$ (i.e. scaffolds are present) results in a large increase in signal duration compared to the case in

which $\beta = 1$. Therefore for a fixed value of τ, the most broadly distributed signal leads to the longest signal duration.

Chapter 3.3 Discussion

We first showed that scaffold concentration is a key variable in regulating the speed of signal transduction. Moreover, we showed that the concentration of a scaffold protein can influence signaling dynamics by controlling the distribution of times over which kinases become active. This type of regulation may have many important consequences that are related to the influence of signal duration on cell decisions. Controlling the times over which kinases are activated may also be useful in directing a specific, robust response in a number of ways. Thus, the scaffold concentration itself provides another variable for maintaining signal specificity by controlling signal duration. This is consistent with data from genetic studies involving KSR1[54, 55], where the authors reported that the concentration of KSR1 can control a cell decision involving commitment to adipogenesis.

Our study focused solely on aspects of scaffold mediated regulation of signal transduction and we only considered the times at which kinases are active in the course of signal transduction. Many other factors also control signal duration. For example, our study does not consider the negative feedback loops that are often associated with the upregulation of phosphatases[59, 72]) or the role of receptor downregulation in controlling signal duration. Also we did not explicitly consider the role of positive versus negative feedback loops in shaping signal duration which is undoubtedly important[70]. It was our focus to study how spatially localizing kinases on a scaffold protein influences signal duration. We aimed to untangle this effect of scaffold proteins from other essential

features of kinase cascades such as allostery and feedback regulation. Also, other theoretical studies have investigated the first passage time statistics in signal transduction cascades and have found interesting dynamics that result from, in part, the sequential activation of multiple steps in a kinase cascade[73, 74]. Our studies of signaling through scaffold proteins supplement these findings and, to our knowledge, provide the first study that shows how scaffolds affect the statistics of signal transduction.

Several predictions from our model of how scaffolds regulate signaling dynamics can be tested. Measurements that monitor the time course of signaling for different scaffold concentrations could potentially resolve the differences in signaling dynamics that are predicted. Also, single molecule or fluorescence correlation based spectroscopic methods[80-82] could potentially probe the statistics of signaling dynamics inherent in kinase cascades and study how such statistics are related to reliable cell decisions. Such techniques can monitor the propagation of a signal, at the level of an individual molecule and thus directly measure how kinase activation within a single cell is distributed over time.

Chapter 3.4. Methods

Chapter 3.4.1 Kinetic Monte Carlo simulations

We simulate a model protein kinase cascade such as the mitogen-activated protein kinase (MAPK) cascade (Fig. 3-1a) in the presence and absence of a scaffold with a kinetic Monte Carlo algorithm[36, 83], which allows us to monitor the relevant stochastic dynamics. Since we are investigating phenomena that occurs on the time scales of signal transduction, we course-grain the system so that proteins are represented as discrete objects, occupying a site on a lattice of dimensions 100 x 100 x 100 lattice spacings.

Scaffold proteins are modeled as rigid, immobile objects containing three binding sites that are each specific for a particular kinase. When bound to a scaffold, kinases are tethered in nearest neighbor positions that are proximal to their downstream substrates. Allowing the scaffold and scaffold-bound species to move does not affect the qualitative results. Reflecting, no flux (i.e.Neumann) boundary conditions exist at each of the faces of the cubic lattice. The system is not periodically replicated since our simulation box is a size on the order a cell. Proteins can diffuse (i.e. translate on the lattice in random directions), bind and unbind, and undergo state transformations according to the prescribed reaction network involving a three staged cascade of activation and deactivation events (Figs. 3-1a, b). Protein motion is subject to excluded volume (steric) constraints in that no two proteins can occupy the same site on the lattice. Chemical (state) transformations and binding events are modeled as thermally activated processes with energy barriers for activation, inactivation, binding and unbinding reactions. Parameters used are given in Table 1.

We simulate the dynamics with a fixed time step Monte Carlo algorithm. In a Monte Carlo step, n trials are attempted, where n is the number of proteins in the simulation. For a given trial, a protein is first chosen at random with uniform probability. A displacement move in a uniformly random direction is attempted with probability,

$$P(diffusion) = \frac{1}{2d} D^{eff} \min\{1, \exp(-E_\infty)\} \qquad (3\text{-}11)$$

where d is the dimensionality of the simulation box, D^{eff} is the probability of attempting a diffusion move and sets an overall time scale to diffuse the length of a lattice site. Excluded volume is accounted for by imposition of an infinite energy barrier, E_∞, for hopping to sites containing other proteins; i.e.

$$E_{\infty} = \begin{Bmatrix} 0 \; ; \; site\,is\,empty \\ \infty \; ; \; site\,is\,occupied \end{Bmatrix} . \qquad (3\text{-}12)$$

Upon considering all possible nearest neighbor interactions, reaction moves, as determined by the network topology, are tried with probabililty,

$$P(reaction) = k^{eff} \min\{1, \exp(-E_{j,j'})\} \qquad (3\text{-}13)$$

where k^{eff} is the probability of attempting a reaction move; (k^{eff} sets an overall reaction time scale), $E_{j,j'}$ is the energy barrier for the $j' \rightarrow j$ reaction scaled with respect to $k_b T$ (Boltzman's thermal energy). With this Monte Carlo move set, the simulations formally evolve the dynamics of the probability $P\left(\vec{r} \mid s^{i,j}; t\right)$ that a chemical species $s^{i,j}$ of type i and state j at position $\vec{r}$ at time t according to the Master equation

$$\begin{aligned} \frac{\partial}{\partial t} P\left(\vec{r} \mid s^{i,j}; t\right) = & \sum_{\vec{r}'} \omega^{i,j}\left(\vec{r} \mid \vec{r}'\right) P\left(\vec{r}' \mid s^{i,j}; t\right) \\ & - \left[\sum_{\vec{r}'} \omega^{i,j}\left(\vec{r}' \mid \vec{r}\right)\right] P\left(\vec{r} \mid s^{i,j}; t\right) \\ & + \sum_{j'} \alpha\left(\vec{r}; j \mid j'\right) \Delta\left(\vec{r} - \vec{r}_k\right) P\left(\vec{r} \mid s^{i,j'}; t\right) \\ & - \left[\sum_{j'} \alpha\left(\vec{r}; j' \mid j\right) \Delta\left(\vec{r} - \vec{r}_k\right)\right] P\left(\vec{r} \mid s^{i,j}; t\right) \\ & + \sum_{\vec{r}'} \sum_{j''} \beta\left(\vec{r} \mid \vec{r}'; j \mid j''\right) \Theta\left(\vec{r}'; i' \mid i''; t\right) P\left(\vec{r} \mid s^{i,j''}; t\right) \\ & - \sum_{\vec{r}'} \sum_{j''} \beta\left(\vec{r} \mid \vec{r}'; j'' \mid j\right) \Theta\left(\vec{r}'; i' \mid i''; t\right) P\left(\vec{r} \mid s^{i,j''}; t\right) \end{aligned} \qquad (3\text{-}14)$$

where $\omega^{i,j}\left(\vec{r}\,|\,\vec{r}\,'\right)$ is the transition probability per unit time for a displacement from $\vec{r}\,'$ to $\vec{r}$ of species $s^{i,j}$; $\omega^{i,j}\left(\vec{r}\,|\,\vec{r}\,'\right)=\frac{D^{eff}}{2d}\min\{1,\exp(-E_{\infty})\}$; $\alpha\left(\vec{r};j\,|\,j'\right)$ is the per unit time transition probability at $\vec{r}$ for a species $s^{i,j'}$ to change to state $s^{i,j}$ (e.g. binding and unbinding reactions) and is $\alpha\left(\vec{r};j\,|\,j'\right)=k^{eff}\min\{1,\exp(-E_{j,j'})\}$, and $\Delta\left(\vec{r}-\vec{r}_k\right)$ imposes the constraint that binding and unbinding occurs only at specified binding sites on the scaffolds at positions $\vec{r}_k$ ($\Delta\left(\vec{r}-\vec{r}_k\right)$ is zero unless a scaffold is located at position $\vec{r}=\vec{r}_k$; and $\beta\left(\vec{r}\,|\,\vec{r}\,';j\,|\,j''\right)=k^{eff}\min\{1,\exp(-E_{j,j''})$ is the transition probability per unit time for a species at $\vec{r}\,'$ to facilitate (i.e. catalyze) the $j''\rightarrow j$ transformation at site $\vec{r}$; and $\Theta\left(\vec{r}\,';i'\,|\,i'';t\right)$ is zero unless the site at $\vec{r}\,'$ is occupied by the appropriate catalyst (i.e. $i'=i''$) in which case it is 1; each summation indicates a sum over nearest neighbors.

Chapter 3.4.2 Parameters used

The parameters used in the simulation were first constrained to typical literature values. Energies of disassociation were taken to be $12k_bT$ corresponding of a disassociation constant K_d of roughly $1\mu M$. 200 stimulatory molecules, S, 200 molecules of kinase A and a 1:1:5 ratio of A, B, and C kinases was used. If we assume a lattice spacing of 10nm, a typical diameter of a protein, the concentration of kinases in our simulation box is roughly $1\mu M$ for kinase A and kinase B and ~5μM for kinase C. In a physiological context, assuming the radius of the cell is about 10μm, this approximately corresponds to

$\sim 10^5$ molecules of kinases A and B and a copy number of $\sim 5 \times 10^5$ for kinase C in our simulation. 600 generic phosphatases are also present. These relative numbers are commensurate with reported kinase concentrations in Yeast and other systems[32, 47]. Chemical kinetics were modeled in the simplest possible way by considering a single elementary reactive collision; i.e.,

$$A^* + B \rightarrow A^* + B^* \qquad (3\text{-}15)$$

where the asterisk (*) denotes an active species. For the purposes of our simulations, saturation effects were ignored and the kinetics were taken to be in a linear regime. Such a model is reasonable when reactions are not limited by the availability of the enzyme. However, relaxing this assumption does not affect the qualitative behavior of our results provided that the times scales involved in the formation of an enzyme-substrate complex and subsequent catalysis do not compete with the diffusive processes in solution. If additional processes associated with enzyme catalysis dominate over diffusive motion of the proteins or binding and unbinding to and from the scaffold, then these process would be observed in the autocorrelation function and corresponding power spectrum. Given that catalysis would incorporate additional processes into the mechanism of kinase activation, such effects would serve to broaden the distribution at all scaffold concentrations as we have observed in our simulations (data not shown). We did not explore this scenario in its entirety since our aim was to solely investigate the effects of scaffolding a kinase cascade.

As discussed in a previous study[12], an important variable that determines the role of scaffolding a kinase cascade is the amount of time required (τ_{ec}) for an active kinase to encounter its downstream target. For simple diffusion, in three

dimensions, $\tau_{ec} \sim \frac{1}{DC^{2/3}}$ where D is the diffusion constant and C is a typical concentration of kinases. Experiments indicate that τ_{ec} is on the order of 10^{-4}s - 10^{0} s[33]. Our studies focused on these experimentally relevant conditions.

Steady-state values are reported. The system is first placed in a random configuration and simulations are allowed to 'equilibrate' by letting the dynamics evolve to a time much larger than the time it takes for a kinase to diffuse the length of the simulation box. Kinases that are inactive at time t' are tagged and waiting times are observed at time $t+t'$ (i.e. statistics are collected for the times at which the kinases become activated), and t' is chosen to be a time longer than the time required for equilibration of the Monte Carlo trajectory.

Chapter 3.4.3 Calculation of statistical quantities

Signaling dynamics can be defined microscopically as the distribution of times at which an individual kinase among of pool of available kinases becomes activated. Therefore, we quantify signaling dynamics by first considering the survival probability $S(t)$. $S(t)$ gives the probability that a particular kinase among the pool of signaling molecules has not been activated at time t provided that it was inactive at time $t=0$. $S(t)$ is a two time point autocorrelation function:

$$S(t)=\langle\sigma(t)\sigma(0)\rangle, \qquad (3\text{-}16)$$

where the brackets denote an ensemble average and $\sigma(t)$ is a binary variable indicating the state of a kinase; i.e.

$$\sigma(t) = \begin{Bmatrix} 0\,; \textit{kinase is active} \\ 1\,; \textit{kinase remains inactive} \end{Bmatrix}. \qquad (3\text{-}17)$$

The survival probability is related to other dynamical properties; for instance, it can be related to a waiting time probability density function or first passage time distribution, $f(t)$, in the following way:

$$S(t) = \int_t^\infty f(t')dt' = 1 - \int_0^t f(t')dt' \qquad and \qquad f(t) = -\frac{d}{dt}S(t). \qquad (3\text{-}18)$$

$S(t)$ is the complement of the cumulative probability distribution of the first passage time. $S(t)$ is computed from the simulations by integrating $f(t)$. Such a calculation is analogous to the data obtained from a single molecule experiment that measures the statistics of enzyme dynamics[80]. This distribution of waiting times underlies the intrinsic duration of signal propagation in a protein kinase cascade—the decay of such a quantity is a measure of how fast the signaling cascade responds to stimuli. Important to note is that this quantity gives information only on the timing of the signal and not on its final magnitude. We also consider the product of the survival probability with the probability that a kinase in the pool of signaling molecules is active in the steady state,

$$R(t) = f_A \langle \sigma(t)\sigma(0) \rangle, \qquad (3\text{-}19)$$

where f_A is the fraction of active kinases at steady state. When normalized, $R(t)$ is a measure of how the activity of the total pool of kinases is distributed over time, and can be thought of as an integrated flux of activated kinases. $-\frac{d}{dt}R(t)$ is seen as a reactive flux in provides a measure of the rate at which downstream kinases are being activated. One can imagine that both quantities could be biologically relevant. If conditions dictate that a biological response requires that a certain number of kinases remain active for extended amounts of time, $R(t)$ may be the more relevant quantity. On the other hand if

the cellular decision requires a count of kinases that become active over a specified time window, then $S(t)$ could be the relevant quantity since it provides a measure of how the activation of individual kinases is distributed over time. Both quantities may be used to integrate signals in different contexts but since our study focuses on signaling dynamics we primarily focus on the survival probability and its related quantities.

Power spectra were computed numerically. Real and imaginary parts of the Fourier transform were obtained from numerical integration using the trapezoidal rule with a step size $\Delta t = 0.001$. $P(\omega)$ is calculated by squaring the real and imaginary parts of $X(\omega)$

$$P(\omega) = \left[\mathrm{Re}\, X(\omega)\right]^2 + \left[\mathrm{Im}\, X(\omega)\right]^2. \qquad (3\text{-}20)$$

$P(\omega)$ was sampled at $N = 100$ logarithmically spaced (i.e. $\omega_{\max} = \omega_0\left(10^{\delta(n-1)}\right)$; $n \in [1,100]$ so that $\delta = \frac{1}{N-1}\log\left[\frac{\omega_{\max}}{\omega_0}\right]$) angular frequencies beginning at: $\omega_0 = \frac{2\pi}{T}$, where T is the total length of the autocorrelation function.

Chapter 3.5 Appendix

Chapter 3.5.1 Time scales for signal transduction

Let us first consider the time scales involved in our model of cell signaling. Kinases in solution must encounter their targeted substrates by diffusion; therefore, encounter (or diffusion) times for the i^{th} kinase, τ_{ec}^{i} (τ_{sig} for the stimulus) are important. These encounter times τ_{ec}^{i}, behave as

$$\tau_{ec}^{i} \sim \frac{1}{D_i \rho_i^{2/3}} \qquad \text{(3-A1)}$$

in three dimensions (cytosolic signaling), for a concentration, ρ_i and diffusivity D_i of each molecular species, i. Other time scales (i.e. τ_k^i, $\tau_p^i, \tau_{on}^i, \tau_{off}^i$) are derived from rates of the many chemical processes taking place in our system. These time scales are for activation (τ_k^i) and deactivation (τ_p^i) by kinases and phosphatases as well as for binding (τ_{on}^i) and unbinding (τ_{off}^i) to and from the scaffold.

Kinases in solution could in principle, upon binding to a scaffold, be assembled into a complex that can not effectively signal. This is because the complex does not have a complete set of kinases bound to it; i.e. the first or second kinase in the sequence is not bound to the scaffold. Such a kinase then would be trapped in a signaling incompetent state until it either disassociates from the complex or the requisite kinases upstream bind to the complex. This scenario requires additional time scales: the times required for a kinase to bind and disassociate from solution into a signaling incompetent complex (τ_{ex3} and τ_{ex4} respectively) and the times required for a signaling incompetent complex to transition to a signaling competent complex (τ_{ex5} and τ_{ex6} respectively) must be accounted for. Processes that determine τ_{ex5} involves upstream kinases binding to the scaffold and is a function of their diffusivities and concentrations; the concentration of scaffolds is also important since that will determine the local concentration of available kinases in solution –thus, $\tau_{ex5} = f(\tau_{ec}^i, \tau_{off}^i, \zeta, \rho_i)$. τ_{ex6} then is the time required for an upstream kinase to unbind from the scaffold, i.e. $\tau_{ex6} = \tau_{off}^i$.

We therefore hypothesized that the source of the observed dynamics in our scaffold-mediated signaling model is a collection of several slower time scales,

$$\tau_{act}^{U}, \tau_{act}^{B}, \tau_{ex1}, \tau_{ex2}, \tau_{ex3}, \tau_{ex4}, \tau_{ex5}, \tau_{ex6}, \qquad \text{(3-A2)}$$

that emerge from the interaction between scaffold density and microscopic reaction and diffusion events (Fig. 4a,b). Let us now investigate the biological consequences of such a minimal scenario that involves the dynamics of competing processes occurring at these eight phenomenological time scales.

Chapter 3.5.2 A Markov model illustrating the competition between the many time scales

We considered a simple course-grained kinetic model in which a kinase K_i can transition between four states denoted with four subscripts: in solution (S), bound to a signaling competent complex (C), bound to an signaling incompetent complex (I), and activated (A). Any bound kinase that is a part of an incomplete complex is said to be in state I. Fig. 3-4a in the main text shows a kinetic graph of the stochastic transitions to neighboring states that involve random waiting times that correspond to a set of eight random variables,

$$\tau_i; \tau_i \in \{\tau_1, \tau_2, ..., \tau_8\}. \qquad \text{(3-A3)}$$

To investigate the simplest scenario, Markovian dynamics are considered. The waiting time for a kinase to transition to a neighboring state is then Poisson distributed with time constants,

$$k_i; k_i \in \{k_1, k_2, ..., k_8\}. \qquad \text{(3-A4)}$$

Thus for the i^{th} process, the waiting time distribution, $F(\tau_i)$, is the first passage time distribution (FPT) and takes the form:

$$F(\tau_i) = k_i e^{-k_i \tau_i} . \tag{3-A5}$$

Ultimately, the quantity of interest is the first-passage time distribution $F(\tau_A)$ for a kinase to transition to its activated state which we denote by $F(\tau_A)$. $F(\tau_A)$ is the time derivative of the cumulative probability distribution (CDF),

$$P(\tau_A < t) \equiv \int_0^t F(\tau')d\tau' ; \tag{3-A6}$$

So that

$$\frac{dP(\tau_A < t)}{dt} = F(\tau_A) . \tag{3-A7}$$

The survival probability is related to the FPT and the CDF in the following way,

$$S(\tau_A) = \int_{\tau_A}^{\infty} F(t')dt' = 1 - \int_0^{\tau_A} F(t')dt' = P(\tau_A < t) . \tag{3-A8}$$

The Master equation is written as follows:

$$\frac{d}{dt}F = Q^+ F - Q^- F \tag{3-A9}$$

where,

$$Q^+ = \begin{pmatrix} 0 & k_7 & k_8 & 0 \\ 0 & 0 & k_6 & k_2 \\ 0 & k_5 & 0 & k_3 \\ 0 & k_1 & k_4 & 0 \end{pmatrix}, \tag{3-A10}$$

and

$$Q^- = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & k_1 + k_5 + k_7 & 0 & 0 \\ 0 & 0 & k_4 + k_6 + k_8 & 0 \\ 0 & 0 & 0 & k_2 + k_3 \end{pmatrix}. \tag{3-A11}$$

So that,

$$\frac{d}{dt}F = QF \;\; ; \;\; Q = Q^{+} - Q^{-} \tag{3-A12}$$

or,

$$\frac{d}{dt}\begin{pmatrix} F_{K_A} \\ F_{K_S} \\ F_{K_C} \\ F_{K_I} \end{pmatrix} = \begin{pmatrix} 0 & k_7 & k_8 & 0 \\ 0 & -(k_1+k_5+k_7) & k_6 & k_2 \\ 0 & k_5 & -(k_4+k_6+k_8) & k_3 \\ 0 & k_1 & k_4 & -(k_2+k_3) \end{pmatrix}\begin{pmatrix} F_{K_A} \\ F_{K_S} \\ F_{K_C} \\ F_{K_I} \end{pmatrix}, \tag{3-A13}$$

with the initial condition,

$$\begin{pmatrix} F_{K_A} \\ F_{K_S} \\ F_{K_C} \\ F_{K_I} \end{pmatrix}_{t=0} = \begin{pmatrix} 0 \\ C_S \\ C_C \\ C_I \end{pmatrix} \tag{3-A14}$$

In principle, an exact solution to the equation can be obtained by finding the eigenvalues, λ_i and eigenvectors $\vec{q}_i$ of the resulting matrix; i.e. the roots of the characteristic polynomial ($\det[Q-\lambda I]=0$) ; however, are solutions to a cubic equation and are too complicated to extract much significant physical information. Therefore, we employed an approximate method that clearly shows the dependence of scaffold density on the behavior of the signaling dynamics. The three cases are considered.

Chapter 3.5.3 Path summation of the Master equation

Formally, we can compute

$$P(\tau_A < t) \tag{3-A15}$$

by considering a weighted sum over all paths that lead to the absorbing state, K_A.

$$P(\tau_A < t) = \sum_{steps,i=1}^{\infty} \sum_{j=1}^{3} C_j \sum_{branches,k=1} \omega\left(\left\{\sum_{jumps,l=1}^{i} \tau_l^{ijk}\right\} < t\right) \prod_{l=1}^{i} \prod_{m} \omega(\tau_l < \tau_{lm}) \tag{3-A16}$$

The first summation decomposes $P(\tau_A < t)$ into separate contributions for each set of paths that contain equivalent numbers of steps required to reach the absorbing state, K_A ; i.e. for i =1, all paths requiring one jump are considered, for i=2, all paths requiring two jumps are considered, etc. Since there can be more than one path containing i steps leading to K_A. The next summation considers the weighted probability that, a priori, a kinase is in one of three states: in solution (S), bound to a signaling incompetent complex (I), and bound to a signaling competent complex (C), i.e. $j \in \{S, I, C\}$ and for normalization,

$$C_S + C_I + C_C = 1 . \qquad \text{(3-A17)}$$

We must then sum over each one of these branches that are denoted with subscript k. A branch is defined here as a particular way in which a path of fixed i and j can be traversed. We account for the probability that a specific path, i, with j steps on the k^{th} branch is taken by computing the probability

$$\omega\left(\left\{\sum_{jumps,l=1}^{i} \tau_l^{ijk}\right\} < t\right) \qquad \text{(3-A18)}$$

that a molecule transitions through a given sequence of jumps, each of which involving a random waiting time τ_l^{ijk}, leading to K_A in a time less than t for the kth branch of the jth molecular state that takes i steps where a path of i steps is composed of individual steps, l . We then avoid overcounting by taking the union of this probability with the joint probability,

$$\prod_{l=1}^{i} \prod_{m} \omega(\tau_l < \tau_{lm}) , \qquad \text{(3-A19)}$$

that no transitions are made in the l^{th} step along the path to any state, m, not along the considered path; τ_l is the waiting time to transition along the l^{th} step of the path and τ_{lm} is the waiting time for a transition at the l^{th} step to a position m that is not along the selected path. This term, $\prod_{l=1}^{i} \prod_{m} \omega(\tau_l < \tau_{lm})$, ensures that each transition l -> l+1 takes place before any transitions to points not along the given path.

Such a path summation is difficult to compute exactly but conveniently lends itself to approximate evaluations. If we note that the contribution of each path requiring i steps, a_i so that

$$P(\tau_A < t) = \sum_{i}^{\infty} a_i \qquad \text{(3-A20)}$$

to the overall cumulative distribution, $P(\tau_A < t)$, decreases monotonically with increasing i, that is,

$$a_{i+1} \le a_i . \qquad \text{(3-A21)}$$

For a normalized probability density function, $f(x)$, i.e. $\int_0^{\infty} f(x)dx = 1$, its cumulative distribution function $P(x < y)$ is less than or equal to 1, $\int_x^{y} P(x')dx' \le 1$.

Therefore, the contribution of each path requiring i steps, a_i (i.e. $P(\tau_A < t) = \sum_{i}^{\infty} a_i$) to the overall cumulative distribution, $P(\tau_A < t)$, decreases monotonically $a_{i+1} \le a_i$ with increasing i; that is,

$$\sum_{j=1}^{3} C_j \sum_{branches,k=1} W\left(\left\{\sum_{jumps,l=1}^{i} \tau_l^{ijk}\right\} < t\right) \prod_{l=1}^{i} \prod_{m} W(\tau_l < \tau_m)$$
$$\geq \sum_{j=1}^{3} C_j \sum_{branches,k=1} W\left(\left\{\sum_{jumps,l=1}^{i+1} \tau_l^{(i+1)jk}\right\} < t\right) \prod_{l=1}^{i+1} \prod_{m} W(\tau_l < \tau_{lm}) \quad \text{(3-A22)}$$

. Thus, the sum can be truncated at all paths requiring i steps with an error that is bounded by $O(a_{i+1})$. As the number of steps, i, increases, the total contribution of each path becomes smaller by a factor involving the ratio of the total contribution to $P(\tau_A < t)$ for the path containing i+1 and i steps; i.e.

$$\frac{a_{i+1}}{a_i} = \frac{\sum_{j=1}^{3} C_j \sum_{branches,k=1} W\left(\left\{\sum_{jumps,l=1}^{i+1} \tau_l^{(i+1)jk}\right\} < t\right) \prod_{l=1}^{i+1} \prod_{m} W(\tau_l < \tau_{lm})}{\sum_{j=1}^{3} C_j \sum_{branches,k=1} W\left(\left\{\sum_{jumps,l=1}^{i} \tau_l^{ijk}\right\} < t\right) \prod_{l=1}^{i} \prod_{m} W(\tau_l < \tau_{lm})} \quad \text{(3-A23)}$$

We can simplify this formula by making use of two identities that hold for Poisson processes. For two independent random variables, τ_i and τ_j, that are exponentially distributed with time constants, k_i and k_j, the probability of τ_i being less than τ_j

$$\omega(\tau_i < \tau_j) = \frac{k_i}{k_i + k_j}. \quad \text{(3-A24)}$$

Also, for a sum of n exponentially distributed random variables

$$\tau_i; \tau_i \in \{\tau_1, \tau_2, ..., \tau_n\} \quad \text{(3-A25)}$$

with time constants,

$$k_i; k_i \in \{k_1, k_2, ..., k_n\}. \quad \text{(3-A26)}$$

The cumulative distribution function (CDF) of the sum of n independent random variables is a convolution of those variables,

$$\omega\left(\sum_i^n \tau_i^{j,k}\right) < t) = \omega\left(\tau_1^{j,k} < t\right) \otimes \ldots \otimes \omega\left(\tau_n^{j,k} < t\right), \qquad \text{(3-A27)}$$

and has the following form:

$$\omega\left(\sum_i^n \tau_i\right) < t) = 1 - \sum_{i=1}^{n}\left(\prod_{j\neq i}\frac{k_j}{k_i - k_j}\right)e^{-k_i t}. \qquad \text{(3-A28)}$$

By substituting these two expressions where appropriate,

$$\frac{a_{i+1}}{a_i} = \frac{\sum_{j=1}^{3} C_j \sum_{branches,k=1} \sum_{n=1}^{i+1}\left(1 - \prod_{j\neq n}\frac{k_j}{k_n - k_j}e^{-k_n t}\right)\prod_{l=1}^{i+1}\prod_m \frac{k_l}{k_l + k_m}}{\sum_{j=1}^{3} C_j \sum_{branches,k=1} \sum_{n=1}^{i}\left(1 - \prod_{j\neq n}\frac{k_j}{k_n - k_j}e^{-k_n t}\right)\prod_{l=1}^{i}\prod_m \frac{k_l}{k_l + k_m}} \qquad \text{(3-A29)}$$

This formula can be rearranged by factoring out the i+1 term inside the summation over the different "branches",

$$\frac{a_{i+1}}{a_i} = \frac{\sum_{j=1}^{3} C_j \sum_{branches,k=1}\left(\left(\left(\sum_{n=1}^{i}\left(1 - \left(\frac{k_{i+1}}{k_n - k_{i+1}}\prod_{j\neq n}\frac{k_j}{k_j - k_n}e^{-k_n t}\right)\right)\right) + \prod_{j\neq i+1}\frac{k_j}{k_j - k_{i+1}}e^{-k_{i+1} t}\right)\prod_{l=1}^{i+1}\prod_m \frac{k_l}{k_l + k_{lm}}\right)}{\sum_{j=1}^{3} C_j \sum_{branches,k=1} \sum_{n=1}^{i}\left(1 - \left(\prod_{j\neq n}\frac{k_j}{k_j - k_n}e^{-k_n t}\right)\right)\prod_{l=1}^{i}\prod_m \frac{k_l}{k_l + k_{lm}}}$$

(3-A30)

Or more compactly,

$$\frac{a_{i+1}}{a_i} = \frac{\sum_{j=1}^{3} C_j \sum_{branches,k=1}\left(\left(\left(\sum_{n=1}^{i} 1 - \beta_n\gamma_n\right) + \alpha\right)\kappa_{i+1}\right)}{\sum_{j=1}^{3} C_j \sum_{branches,k=1} \sum_{n=1}^{i}(1 - \gamma_n)\kappa_i} \qquad \text{(3-A31)}$$

where,

$$\alpha = \prod_{j\neq i+1}\frac{k_j}{k_j - k_{i+1}}e^{-k_{i+1} t},$$

$$\beta_n = \frac{k_{i+1}}{k_{i+1} - k_n},$$

$$\gamma_n = \prod_{j \neq i+1} \frac{k_j}{k_j - k_n} e^{-k_n t},$$

and

$$\kappa_i = \prod_{l=1}^{i} \prod_{m} \frac{k_l}{k_l + k_{lm}}.$$

Thus the error, E, that is introduced by truncating the sum at a given number of steps is

$$E = O\left(\frac{\sum_{j=1}^{3} C_j \sum_{branches,k=1} \left(\left(\left(\sum_{n=1}^{i} 1 - \beta_n \gamma_n \right) + \alpha \right) \kappa_{i+1} \right)}{\sum_{j=1}^{3} C_j \sum_{branches,k=1} \sum_{n=1}^{i} (1 - \gamma_n) \kappa_i} \right) \quad \text{(3-A32)}$$

From the formula (3-A32), we see that many conditions allow for

$$\frac{a_{i+1}}{a_i} << 1 \quad \text{(3-A33)}$$

in which case, the summation quickly decays and can be truncated at i steps. Moreover, any significant difference in time scales for processes in successive steps results in such a decay. Now we consider the application of this model to the situations that involve different scaffold expression levels. Summing from j=1 to j=3 gives us

$$P(\tau_A < t) = P(\tau_{A,solution} < t) + P(\tau_{A,incomplete} < t) + P(\tau_{A,complete} < t) \quad \text{(3-A34)}$$

where $P(\tau_A < t)$. This gives us the cumulative density as a composition of many terms contributing from initial states; from solution, incomplete complexes, and complete complexes where,

$$P(\tau_{A,solution} < t) = C_s \left(P(\tau_7 < t) P(\tau_7 < \tau_1) P(\tau_7 < \tau_5) \right) +$$
$$\left(P(\tau_1 + \tau_2 + \tau_7 < t) P(\tau_1 < \tau_5) P(\tau_1 < \tau_7) P(\tau_2 < \tau_3) P(\tau_7 < \tau_1) P(\tau_7 < \tau_5) \right)$$

$$P(\tau_{A,incomplete} < t) = C_I \left(P(\tau_2 + \tau_7 < t) P(\tau_2 < \tau_3) P(\tau_7 < \tau_1) P(\tau_7 < \tau_5) \right) +$$
$$P(\tau_3 + \tau_8 < t) P(\tau_3 < \tau_2) P(\tau_8 < \tau_4) P(\tau_8 < \tau_6) +$$
$$\left(P(\tau_2 + \tau_5 + \tau_8 < t) P(\tau_2 < \tau_3) P(\tau_5 < \tau_1) P(\tau_5 < \tau_7) P(\tau_8 < \tau_4) P(\tau_8 < \tau_6) \right) +$$
$$\left(P(\tau_3 + \tau_6 + \tau_7 < t) P(\tau_3 < \tau_2) P(\tau_6 < \tau_4) P(\tau_6 < \tau_8) P(\tau_7 < \tau_1) P(\tau_7 < \tau_3) \right)$$

$$P(\tau_{A,complete} < t) = C_C\left(P(\tau_8 < t)P(\tau_8 < \tau_4)P(\tau_8 < \tau_6)\right) +$$
$$P(\tau_6 + \tau_7 < t)P(\tau_6 < \tau_4)P(\tau_6 < \tau_8)P(\tau_8 < \tau_6)P(\tau_7 < \tau_1)P(\tau_7 < \tau_5) +$$
$$\left(P(\tau_3 + \tau_4 + \tau_8 < t)P(\tau_4 < \tau_6)P(\tau_3 < \tau_2)P(\tau_4 < \tau_8)P(\tau_8 < \tau_4)P(\tau_8 < \tau_6)\right)$$

First, consider the case in which scaffold proteins are present in negligible amounts and signaling is distributed over one exponential time scale ($\beta \sim 1.0$). What we aim to show is that a separation of time scales is apparent and that the dominant time scale is the origin of the exponential decay. For low scaffold concentrations ($C_S \sim 1$), most kinases are present in solution and the transition to either scaffold-bound state is very slow, i.e. $k_7 >> k_5$ and $k_7 >> k_1$, the summation can therefore be cut off at one step giving:

$$P(\tau_A < t) = 1 - \left(\frac{k_7}{k_7 + k_1}\right)\left(\frac{k_7}{k_7 + k_5}\right)e^{-k_7 t} + h.o. \approx 1 - e^{-k_7 t}. \qquad \text{(3-A35)}$$

If signaling was only allowed to take part on a scaffold (i.e. $k_7 = 0$) then by the same argument,

$$P(\tau_A < t) \approx 1 - e^{-k_5 t}$$

As expected, single exponential behavior in the first passage time is a signature of one dominant time scale controlling the signaling process.

Now consider the situation where scaffold concentration is very high (i.e. $C_I \approx 1$) and again, $\beta \sim 1.0$ suggesting the presence of a dominant time-scale. In this case we consider paths starting from K_I that end in the active state. Starting at K_I, there are two branches that lead to

$$K_A,\ K_I \xrightarrow{2} K_S \xrightarrow{7} K_A \qquad \text{(3-A36)}$$

and

$$K_I \xrightarrow{3} K_C \xrightarrow{8} K_A. \qquad \text{(3-A37)}$$

However, the dominant time scale in this scenario is k_3^{-1}. Keeping only the branch that involves process 3 we have

$$\begin{aligned} P(\tau_A < t) &= P(\tau_3 + \tau_8 < t)P(\tau_3 < \tau_4)P(\tau_8 < \tau_4)P(\tau_8 < \tau_6) \\ &= 1 - \left(\frac{k_3}{k_3 + k_4}\right)\left(\frac{k_8}{k_8 + k_4}\right)\left(\frac{k_8}{k_8 + k_6}\right)\left(\frac{k_8}{k_8 - k_3}e^{-k_3 t} + \frac{k_3}{k_3 - k_8}e^{-k_8 t}\right) + h.o. \qquad \text{(3-A38)} \\ &\approx 1 - \left(\frac{k_3 k_8^2}{k_4\left(k_8 + k_4\right)\left(k_8 + k_6\right)}\right)e^{-k_3 t} \end{aligned}$$

Since $k_4 >> k_3$ and $k_8 >> k_3$.

In the intermediate regime of scaffold density, no such separation of time scales is apparent; also, there is significant probability that a kinase resides in any of the three initial states. In our simulations, binding affinity, K_d, was taken to be ~1um (1kT/molecule ~ 0.6 kCal/mol) and this corresponds to a binding energy

$$F = -k_b T \ln\left(K_d\right) \qquad \text{(3-A39)}$$

of roughly 12-15kT. Given this binding affinity and these kinase concentrations, roughly 85% of MAP3K and MAP2K are bound. This implies that ~72% of the bound kinase exist in fully assembled signaling-competent complexes. The number in solution is the number of MAPK in excess solution and 5x as many MAPK as scaffolds were used in the simulation. Therefore, $K_s \sim 0.8$, $K_I \sim 0.06$, $K_s \sim 0.14$. Since all eight time-scales are relevant, the summation requires more terms;

$$P(\tau_A < t) = P(\tau_{A,solution} < t) + P(\tau_{A,incomplete} < t) + P(\tau_{A,complete} < t) \qquad \text{(3-A40)}$$

where,

$$P(\tau_A < t) = 1 - S(t).$$

This expression gives us, when decomposed into a contribution from solution, incomplete complexes, and complete complexes, a superposition of many exponential terms. In this case one can see from the formula that the cumulative distribution is a composition of many characteristic time scales that govern signaling dynamics.

More complexity is added to this mechanism if the transitions between different signaling states are more complicated than simple Poisson processes (e.g. memory effects are important). This may very well be the case since many nonlinear processes are taking part in each fundamental event in our mesoscopic model and are not necessarily uncorrelated in time. We showed that a multi-state kinetic model with Markovian dynamics can give rise to multi-exponential kinetics in some range of scaffold concentrations. It is likely, however, that transitions between these states in actuality have more complicated transitions. For example binding events in the model are related to the first-passage time distribution for a kinase to encounter a scaffold with an available binding pocket. In the continuum limit, the first-passage time for a random walker in one dimension with diffusion constant, D, to travel a distance, x_o, for the first time, t has the form:

$$F(x_o;t) = \frac{x_o}{(4\pi D t)^{1/2}} \frac{e^{-x_o^2/4Dt}}{t}. \qquad \text{(3-A41)}$$

The purpose of the Markov model then was to show that complex, multi-exponential behavior in signaling dynamics can arise from the mixing of many different routes through which a signal can be propagated.

Chapter 4 – Signal Duration and the Time Scale Dependence of Signal Integration in Biochemical Pathways

Chapter 4.1. Introduction

Signal duration (e.g. the length of time over which a signaling intermediate is active) is a critical determinant in mediating cell decisions in numerous biological processes(Fig. 1) including cell growth, proliferation, and developmental lineage commitments [58-62, 64, 66, 70]. One fundamental issue in signal transduction and cell decision making then is how differences in signal duration are detected to achieve the appropriate biological response.

Accompanying changes in signal duration are multiple stages of biochemical regulation of differing network topology that collectively integrate an incoming signal to deliver a specific biological response. The sequential activation of multiple steps in a biochemical pathway is a ubiquitous regulatory motif involved in many aspects of gene regulation, metabolism, and intracellular signal transduction. Many advantages of having multiple steps of regulation as opposed to having activation occur through a single step have been documented. A signaling cascade can allow for attenuation of noise, incorporation of additional regulatory checkpoints or proofreading steps, and increased tunability of the input signal[21, 22, 84, 85]. Other studies have also established conditions under which signaling cascades amplify or attenuate incoming signals[12, 31, 86]. These conditions are established by rates of activation, rates of deactivation that are set by phosphatase activities, and the presence of scaffold proteins[12, 13, 31]. However, how these features synergize to detect differences in signal duration has not been fully studied.

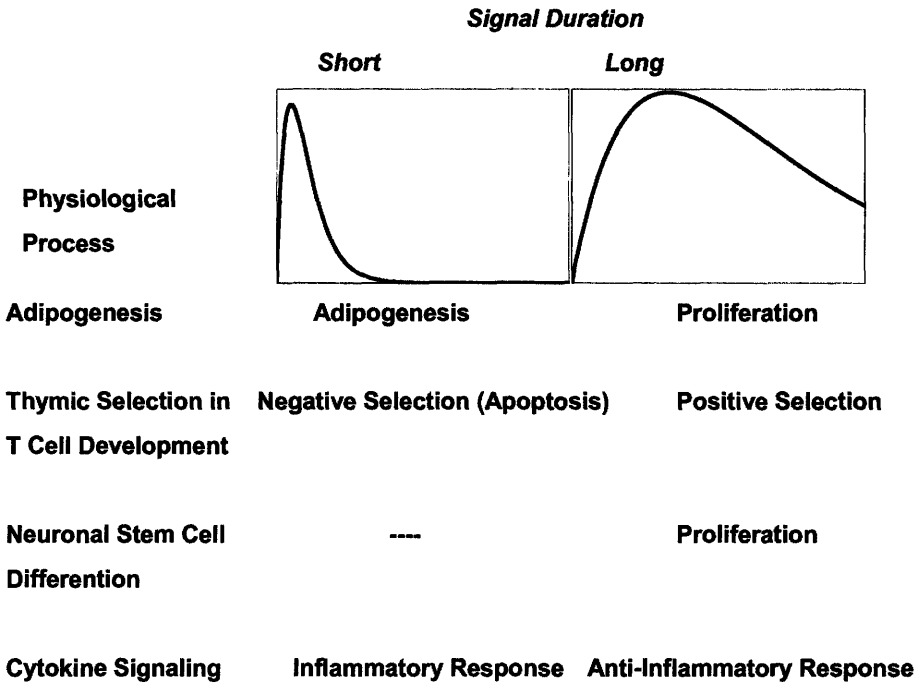


Figure 1. **Physiological examples of signal duration determining the phenotypic outcome in signal transduction**

Four examples of branching physiological processes in which phenotypic decisions are believed to be controlled by the detection of differences in signal duration[55, 58, 60, 65, 70, 87].

A recent study has proposed a model that predicts how signals with different dynamical characteristics can be distinguished upon integration into network architectures[88, 89]. We develop a formalism to complement their approach and, as a consequence, identify general principles for how different network topologies can differentially integrate signals that persist at different time scales. We focus on a simple model of the sequential enzymatic activation of multiple species along a pathway to understand mechanistic principles underlying how multiple stages in a biochemical pathway can integrate differences in signal duration. We use a model of a weakly

activated cascade[31, 86, 90], whose assumptions we first motivate, to investigate general principles underlying how biochemical cascades detect the time scale dependence of input signals. The model allows us to characterize the dynamics by obtaining exact expressions for the power spectra of linearized biochemical networks of multiple stages with arbitrary length and connectivity.

Our analysis emphasizes how signals that persist at different time scales multiple staged biochemical pathway. We first show that biochemical cascades can function as both high low and high pass filters depending on the topology of the network architecture. A low pass filter removes high frequency (short duration) components of a signal and a high pass filter removes low frequency (long duration) components of a signal. These filtering capabilities are determined by differential positive and negative regulation within the biochemical module that occurs at different time scales. Importantly, the filtering capabilities are determined by the presence of feedback as well as the amplification and attenuation properties at different steps in the cascade that are set by the differences in phosphatase activities at different stages along the cascade. Our findings suggest design principles that characterize how biochemical cascades are well suited for detecting time scale dependent differences in biochemical signals.

Chapter 4.2. Results

In the model, an input signal, $f(t)$, activates the first member of the pathway whose activation can then activate the next member. In turn, each upstream species activates its immediate downstream target and can also be deactivated by, for example, a phosphatase. Assuming Michaelis Menten kinetics for the activation and deactivation of

each species along the cascade, we can write an equation for the dynamics of the active form of the i^{th} species x_i along the cascade:

$$\frac{dx_i}{dt} = \frac{k^i_{cat}\, x_{i-1}\left(x_i^T - x_i\right)}{K^i_M + \left(x_i^T - x_i\right)} - \frac{k^i_{cat,pase} E_{pase}\, x_i}{K^i_{M,pase} + x_i}, \qquad (4\text{-}1)$$

where E_{pase} is the concentration of the enzyme that deactivates species i, $k^i_{cat}, k^i_{cat,pase}$ are the catalytic constants of the activation and deactivation steps, $k^i_M, k^i_{M,pase}$ are the Michaelis constants and x_i^T is the total amount species available at step i. We take $K^i_M, K^i_{M,pase}$ at each step to be large ($K^i_M, K^i_{M,pase} > x_i^T$) so that the kinetics of the reactions are not limited by the availability of the enzyme[34]. Next, if we assume that the cascade is weakly activated (i.e. at each stage, the total number of species is much larger than the number of active species, $x_i^T >> x_i$). In many biologically relevant instances (e.g. the Mitogen activated protein kinase (MAPK) cascade), the neglect of saturation effects is reasonable[91]. Further, modeling the deactivation as a first order reaction is often valid when phosphatases are in excess as is the case in many physiological scenarios[47]. Eq. 1 simplifies to a system of linear first order differential equations[31]:

$$\begin{aligned} \frac{dx_1}{dt} &= k_1^+ f(t) - k_1^- x_1 \\ \frac{dx_i}{dt} &= k_i^+ x_{i-1} - k_i^- x_i \end{aligned}, \qquad (4\text{-}2)$$

where the first species is activated at a rate $k_1^+ f(t)$; $k_i^+ = \frac{x_i^T k^i_{cat}}{K_{M,i}}$ and $k_i^+ = \frac{E_{pase} k^i_{cat,pase}}{K^i_{M,pase}}$.

The weakly activated cascade model has the advantage that the linearity of the equations allows for analytical tractability. Eq. 4-2 can be conveniently analyzed by introducing Fourier transformed variables: $X_i(\omega) = \int dt e^{i\omega t} x_i(t)$ and $F(\omega) = \int dt e^{i\omega t} f(t)$. The number of activated species at stage i becomes $X_i(\omega) = \frac{k_i^+ X_{i-1}(\omega)}{i\omega + k_i^-}$. The power spectrum $P_i(\omega) \equiv |X_i(\omega)|^2$ at the i^{th} step can be obtained $P_i(\omega) = \frac{(k_i^+)^2}{\omega^2 + (k_i^-)^2} P_{i-1}(\omega)$. After iterating at each successive stage of the n step cascade, an expression for $P_n(\omega)$ as a function of the power spectrum of the input signal ($S(\omega) \equiv |F(\omega)|^2$) is obtained:

$$P_n(\omega) = g_n(\omega) S(\omega), \tag{4-3}$$

in which a frequency dependent gain $g_n(\omega)$ is defined as:

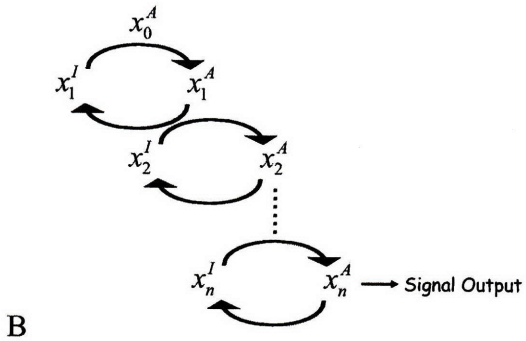
$$g_n(\omega) = \prod_{i=1}^{n} \left(\frac{k_i^+}{k_i^-} \right)^2 \frac{1}{1 + \left(\frac{\omega}{k_i^-} \right)^2}. \tag{4-4}$$

$g_n(\omega)$ is a transfer function that converts the input $S(\omega)$ into a response and provides a measure of the signal processing capabilities of the network. The change in the amplitude of the signal output is determined by the $\prod_{i=1}^{n} \left(\frac{k_i^+}{k_i^-} \right)^2$ term and the time scale dependence of the output is modulated by the $\prod_{i=1}^{n} \left[1 + \left(\frac{\omega}{k_i^-} \right)^2 \right]^{-1}$ term.

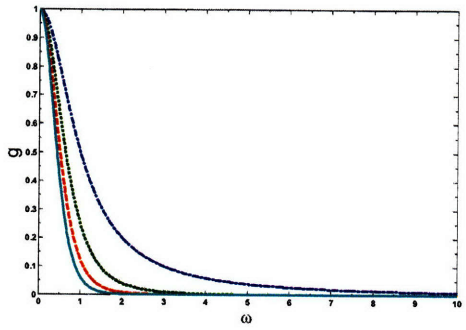
From the formula of $g_n(\omega)$, one consequence of having multiple stages is readily apparent. In the high frequency regime $\left(\frac{\omega}{k_i^-} >> 1\right)$, for each k_i^-, $g_n(\omega)$ rapidly decays with increasing n ($g_n(\omega) \sim \omega^{-2n}$). Thus, longer cascades are more efficient at filtering the high frequency components of the signal from the output. This behavior is illustrated in Fig. 4-2b. Fig. 4-2b contains plots of $g_n(\omega)$ for different values of n; cascades of lengths $n = 1, 2, 3, 4$ are shown.

In eq. 4-4, the relative values of k_i^- along different stages of the cascade also affects the scaling behavior of $g_n(\omega)$ as ω changes as well as the overall amplitude. The change in signal amplitude at the steady state (that leads to amplification or attenuation) at step i is given by the ratio of the effective rate constants for activation and deactivation $\frac{k_i^+}{k_i^-}$. $\frac{k_i^+}{k_i^-} > 1$ results in signal amplification and $\frac{k_i^+}{k_i^-} < 1$ leads to attenuation of the signal at step i [31, 86]. Amplification or attenuation also leads to different frequency dependent behaviors of $g_n(\omega)$. For example, consider an n staged cascade with rate constants k_i^- such that $k_n^- > k_{n-1}^- > k_{n-2}^- ... k_2^- > k_1^-$. At frequencies $\omega > k_n^-$, $g_n(\omega) \sim \omega^{-2n}$ while at frequencies $k_n^- > \omega > k_{n-1}^-$, $g_n(\omega) \sim \omega^{-2(n-1)}$ and so forth. Thus, at intermediate frequencies, a time scale separation (as determined by different deactivation rates along the cascade), along with signal amplification and attenuation, also leads to different frequency dependent behaviors of $g_n(\omega)$.

A



B



C

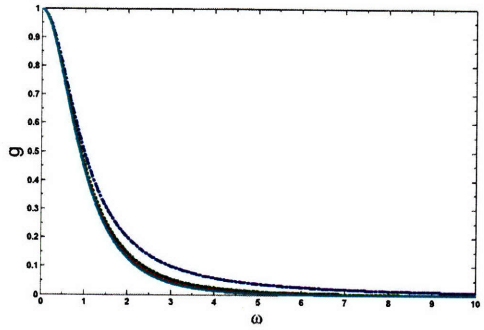


Figure 4-2. **Filtering of high frequency signals**

Time scale dependence of signal integration in a biochemical cascade. a.) the sequential activation of multiple stages in a signaling cascade. Superscripts (I) and (A) denote inactive and active forms of each chemical species and are dropped from the equations in the text. b.) same kinetic constants, all kinetic constants are taken to be: $k_i^+ = k_i^- = 1.0$ c.) a positive gradient

activation/deactivation rates keeping $\frac{k_i^+}{k_i^-} = 1.0$ fixed. $k_1^+ = 1.0$, $k_2^+ = 3.3$, $k_3^+ = 6.6$, $k_4^+ = 10.0$

Fig. 4-2c contains plots of $g_n(\omega)$ for $n = 1,2,3,4$ with successively different values of k_i^- while keeping $\frac{k_i^+}{k_i^-} = 1.0$ fixed ($k_1^- = 1.0$, $k_2^- = 3.3$, $k_3^- = 6.6$, $k_4^- = 10.0$).

Also, from the plots in Fig. 2c (and inspection of eq. 4-4), it is observed that incorporation of faster steps along the cascade influences the frequency dependence of $g_n(\omega)$ to a lesser extent than would be case when kinetics of activation are same for each successive step. Since signal propagation in these cases is limited by the slower stages of the cascade, the faster steps are effectively removed from $g_n(\omega)$. This observation suggests a design principle in the ability of a biochemical pathway to filter signals of short duration: when there is a positive gradient of deactivation rates (that also leads to amplification or attenuation of signal amplitude), the time scale dependence of signal integration for multi staged cascades more closely resembles that of a pathway involving a single step.

While the sequential activation of multiple steps in a biochemical pathway allows for effective filtering of the high frequency components of a signal, often the desired signal output is regulated by feedback. We will now show that feedback control in some instances also allows for the filtering of the low frequency (long time) components of an active signal. In these instances, signals that occur at short times can be integrated while signals with a longer duration are effectively filtered.

For instance, the signal output can be affected by a feedback loop that is initiated downstream of the output. This scenario would be the case when the signal output from a biochemical cascade activates its own positive or negative regulators. For instance, in mammalian cells, the activation of extra cellular regulatory kinase (ERK) often leads to the upregulation or activation of its own phosphatases[92]. In this scenario, signal output is activity of the kinase at the m^{th} step and feedback control to the signal output at step m is initiated at a later step (i.e. $n > m$) then the modified set of dynamical equations becomes:

$$\begin{aligned} \frac{dx_1}{dt} &= k_1^+ f(t) - k_1^- x_1 \\ \frac{dx_m}{dt} &= k_m^+ x_{m-1} - k_m^- x_m + \upsilon\, k^f x_n \end{aligned} \tag{4-5}$$

where k^f is the feedback strength and sets the time scale of the feedback and $\upsilon = \left\{ \begin{matrix} 1\,; positive\ feedback \\ -1\,; negative\ feedback \end{matrix} \right\}$. After applying a Fourier transformation as before, an expression for $P_m^f(\omega)$, albeit now more complicated, can be obtained as a function of $S(\omega)$ in closed-form:

$$P_m^f(\omega) = g_m^f(\omega) S(\omega), \tag{4-6}$$

where,

$$g_m^f(\omega) = \prod_{i=1}^{m} \frac{\left(k_i^+\right)^2}{\left(\left(k_i^-\right)^2 + \omega^2\right)} + \frac{\left(k^f\right)^2\left[\prod_{i=1}^{n}\left(k_i^+\right)^2\right]}{\left(\left(k_m^-\right)^2 + \omega^2\right)\prod_{i=1}^{n}\left(\left(k_i^-\right)^2 + \omega^2\right) + \left(k^f\right)^2 \prod_{i=m+1}^{n}\left(k_i^+\right)^2 \prod_{i=1}^{m-1}\left(\left(k_i^-\right)^2 + \omega^2\right) + 2\prod_{i=m}^{n}\left(k_i^+\right)^2 \prod_{i=1}^{m-1}\left(\left(k_i^-\right)^2 + \omega^2\right)} + \frac{k^f\left[\prod_{i=1}^{n}\left(k_i^+\right)\right]\left[\prod_{i=1}^{m}\left(k_i^+\right)\right]}{\frac{\upsilon}{2}\left[\prod_{i=m}^{n}\left(k_i^-\right)\right]\prod_{i=1}^{m}\left(\left(k_i^-\right)^2 + \omega^2\right) - \upsilon k^f\left(\left(k_m^-\right)^2 + \omega^2\right)\prod_{i=m+1}^{n}\left(k_i^+\right)\prod_{i=1}^{m-1}\left(\left(k_i^-\right)^2 + \omega^2\right)}$$

(4-7).

In the case of feedback regulation, there exists a competition between processes that are realized on multiple time scales: one for the signal to propagate along the cascade to the species involved in the signal output, and the others for the additional interactions derived from the feedback loops to propagate and interact with the species involved in the output. The competition between these effects in principle may lead to a frequency dependent optimal value of $g_m^f(\omega)$. At high frequencies as before, signal propagation is limited by the time it takes to move through the cascade and high frequency components of $g_m^f(\omega)$ are filtered. Also, at low frequencies, signals can potentially be attenuated when the response is dominated by the activity of the feedback loop. If the interaction from the feedback is sufficiently strong, then the low frequency components of the signal are also filtered by the cascade. In this scenario, the frequency dependent behavior of $g_m^f(\omega)$ would be non monotonic.

We illustrate these ideas through consideration of a three tiered cascade that consists of a chemical species carrying the input signal, a species conferring the signal

output, and a species activated downstream to the output that provides a feedback interaction to the species conferring the signal output. In this scheme, $m=1$ and $n=2$, and eq. 4-7 is simplified and $g_1^f(\omega)$ becomes:

$$g_1^f(\omega)=\frac{\left(k_1^+\right)^2\left[\left(k_2^-\right)^2+\omega^2\right]}{\left(k_1^-k_2^--\upsilon k_f k_2^+\right)^2+\omega^2\left(\left(k_1^-\right)^2+\left(k_2^-\right)^2+2\upsilon k_f k_2^+ +\ \omega^2\right)}. \qquad (4\text{-}8)$$

The optimal frequency ω_{opt} is obtained by differentiating $g^f{}_1(\omega)$,

$$\omega_{opt}=\left[\left\{k_f k_2^+\left(k_f k_2^+-2\upsilon k_2^-\left(k_1^-+k_2^-\right)\right)\right\}^{1/2}-\left(k_2^-\right)^2\right]^{1/2}. \qquad (4\text{-}9)$$

ω_{opt} increases monotonically for decreasing values of k_2^- and increasing values of k_f. For negative feedback $\upsilon=-1$, ω_{opt} exists ($\mathrm{Im}\,\omega_{opt}=0$) when $\left[k_f k_2^+\left\{k_f k_2^++2k_2^-\left(k_1^-+k_2^-\right)\right\}\right]^{1/4}>k_2^-$. Positive feedback $\upsilon=+1$ requires that two conditions are satisfied for ω_{opt} to be real, $\left[k_f k_2^+\left(k_f k_2^+-2k_1^-k_2^-\right)\right]^{1/4}>k_2^-$ and $k_f k_2^+>2k_2^-\left(k_1^-+k_2^-\right)$. The height at the optimal frequency $g_1^f\left(\omega_{opt}\right)$ is:

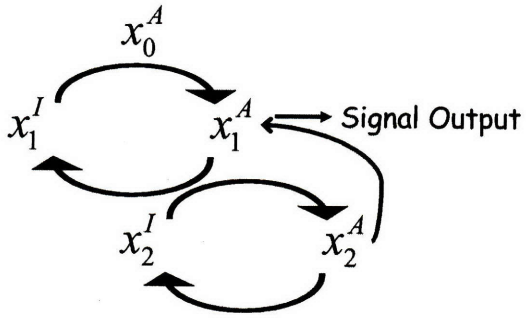
$$g_1^f\left(\omega_{opt}\right)=\frac{\left(k_1^+\right)^2}{\left(k_1^-\right)^2-\left(k_1^-\right)^2+2\left\{k_f k_2^++\sqrt{k_f k_2^+\left[k_f k_2^+-2k_2^-\left(k_1^-+k_2^-\right)\right]}\right\}}. \qquad (4\text{-}10)$$

We can also compute the width $\omega_{1/2}$ of $g_1^f(\omega)$ at half maximum $g_1^f\left(\omega_{opt}\right)/2$. $\omega_{1/2}$ has the form: $\omega_{1/2}=\frac{1}{\sqrt{2}}\left[\sqrt{4\gamma_1+\gamma_2+\sqrt{8\gamma_1\left(2\gamma_1+2\gamma_3\right)+\gamma_4}}-\sqrt{4\gamma_1+\gamma_2-\sqrt{8\gamma_1\left(2\gamma_1+2\gamma_3\right)+\gamma_4}}\right]$,

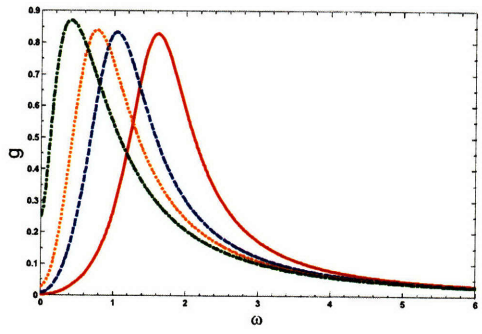
(11)

where $\gamma_1 = \sqrt{k_f k_2^+ \left[k_f k_2^+ - 2k_2^- \left(k_1^- + k_2^- \right) \right]}$, $\gamma_2 = \left(k_1^- \right)^2 - 3\left(k_2^- \right)^2 + 2k_f k_2^+$, $\gamma_3 = \gamma_2 + 2\left(k_2^- \right)^2$, and $\gamma_4 = \left(k_1^- + k_2^- \right)^2 \left(k_1^- - k_2^- \right)^2 + 4k_f k_2^+$. Fig. 4-3b considers plots of $g_1^f(\omega)$ for different feedback strengths k_f. The curves in Fig. 3b illustrate changes in $g_1^f(\omega)$, $g_1^f(\omega_{opt})$, and $\omega_{1/2}$. Fig. 4-3c illustrates how $g_1^f(\omega)$, $g_1^f(\omega_{opt})$, and $\omega_{1/2}$ change for different values of k_2^-.

A



B



C

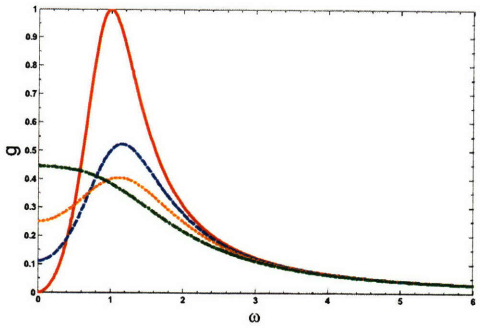


Figure 4-3. **Filtering of low frequency signals**

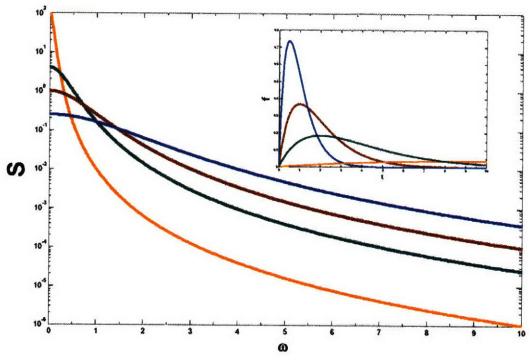
A three tiered biochemical cascade with competing processes occurring at different time scales is sufficient to filter signals at long time scales (Low pass filtering). a.) An initial stimulus activates a downstream species that confers a signal output and activates a downstream species that, through feedback, interacts with the species that carries the signal output. Superscripts (I) and (A) denote inactive and active forms of each chemical species and are dropped from the equations in the text. b.) For each plot, these parameter values were taken to be: $\upsilon = -1$, $k_1^+ = k_1^- = k_2^+ = 1.0$; $k_2^- = 0.1$. $k_f = 2.5$ (solid lines), $k_f = 1.0$ (dashed lines), $k_f = 0.5$ (dotted lines), $k_f = 0.1$ (dash-dotted lines). c.) parameter values were taken to be: $\upsilon = -1$, $k_1^+ = k_1^- = k_2^+ = k_f = 1.0$. $k_2^- = 0.0$ (solid lines), $k_2^- = 0.5$ (dashed lines), $k_2^- = 1.0$ (dotted lines), $k_2^- = 2.0$ (dash-dotted lines).

A convenient way to parameterize signals of differing duration, while keeping the total amount of signal $\int f(t)\,dt = \alpha$ fixed, is to consider the function

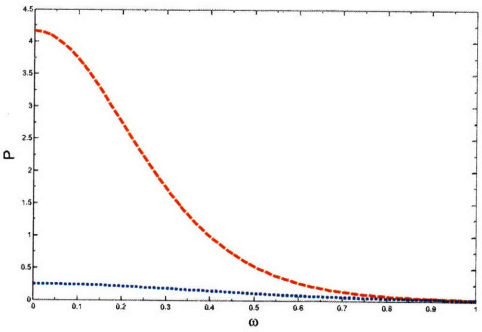
$f(t) = \alpha \dfrac{te^{-t/\tau_d}\Theta(t)}{\int te^{-t/\tau_d}\Theta(t)\,dt}$, where $\Theta(t)$ is a Heaviside step function, τ_d sets the signal duration, and α is taken to be 1 ($\alpha = 1$) in the appropriate units. This form of $f(t)$ models the behavior of a typical experimental signaling time course[59]. For this choice of signal, $S(\omega)$ is easily computed; $S(\omega) = \dfrac{\alpha^2}{\left(\tau_d^{-2} + \omega^2\right)^2}$. $S(\omega)$ is plotted in Fig. 4a for

different values of τ_d ranging from $\tau_d^{-1} = 2.0$ (short duration) to $\tau_d^{-1} = 0.1$ (long duration). The corresponding plots of $f(t)$ are shown on the inset of Fig. 4-4a.

A



B



C

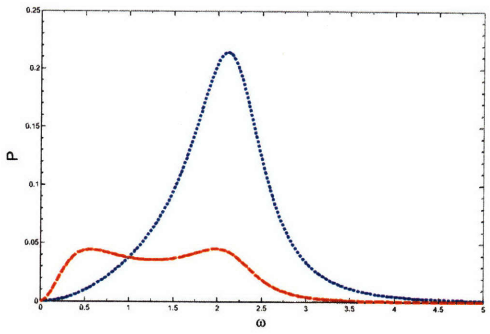


Figure 4. **Integration of differences in signal duration**

a.) Differences in signal duration parameterized by τ_{deg}; $\tau_{\text{deg}}^{-1} = 0.1$ (dash-dotted), 0.5 (dotted), 1.0 (dashed), 2.0 (solid) lines. Plots of $S(\omega) \equiv |F(\omega)|^2$ are shown. Corresponding plots of $f(t)$ are shown in the inset. Short $\tau_{\text{deg}}^{-1} = 2.0$ and long $\tau_{\text{deg}}^{-1} = 0.5$ duration signals are filtered through b.) $g_3(\omega)$ and c.) $g_1^f(\omega)$ resulting in b.) $P_3(\omega)$ and c.) $P_1^f(\omega)$ for $\tau_{\text{deg}}^{-1} = 2.0$ (dashed lines) and $\tau_{\text{deg}}^{-1} = 0.5$ (dotted lines). Parameters taken to be: b.) c) $k_1^+ = 2.0$, $k_2^+ = 1.0$, $k_1^- = 1.0$, $k_2^- = 0.01$, $k^f = -5.0$.

Figs. 4-4b,c illustrates how signals $S(\omega)$ of large ($\tau_{\text{d}}^{-1} = 0.5$ dotted lines) and small ($\tau_{\text{d}}^{-1} = 2.0$ dashed lines) duration are integrated by the internal gains $g_3(\omega)$ and $g_1^f(\omega)$ of these multistage cascades of differing network topologies. In Fig. 4-4b, the signal output $P_3(\omega)$, upon integration by a three-tiered kinase cascade is shown. Taking $k_i^+ = k_i^- = 1.0$ for $i \in 1,2,3$, $g_3(\omega)$ effectively filters the short ($\tau_{\text{d}}^{-1} = 2.0$) duration signal and results in an output $P_3(\omega)$ of small magnitude at all time scales $2\pi\omega^{-1}$ in the frequency spectrum. In contrast, for the signal characterized by $\tau_{\text{d}}^{-1} = 0.5$, signal processing through $g_3(\omega)$ results in a signal of larger amplitude. The ratio of amplitudes

$\frac{P_3^{\tau_d=0.5}}{P_3^{\tau_d=2.0}}$ at the optimal frequency mode ($\omega = 0$) for the two signals is $\frac{P_3^{\tau_d=0.5}}{P_3^{\tau_d=2.0}} \approx 17$ In Fig. 4c, the signal output $P_1^f(\omega)$, obtained from a signal output that is also affected by a downstream negative ($\upsilon = -1$) feedback loop, is shown. Parameters used are: $k_1^+ = 2.0$, $k_2^+ = 1.0$, $k_1^- = 1.0$, $k_2^- = 0.01$, $k^f = -5.0$. For the signal of long duration $\tau_d^{-1} = 0.5$, only the small frequency components of the signal are integrated. This behavior is in contrast with the signal output of a short duration signal $\tau_d^{-1} = 2.0$. The amplitude difference in this case is $\frac{P_1^{f,\tau_d=0.5}}{P_1^{f,\tau_d=2.0}} \approx 0.2$.

Chapter 4.3. Conclusions

We demonstrated how simple network structures involving multiple steps of biochemical regulation can differentially detect signals that have different temporal behaviors. Our models illustrate the topological requirements needed for different network motifs to detect signals of differing duration. To illustrate these effects, we computed the frequency dependent internal gain for two classes of biochemical pathways involving multiple stages of regulation. The first model consisted of a cascade of steps and showed how changes in the number of steps as well as the amplification/attention changed the networks' ability to filter high frequency (short duration) components of a signal. The second model consisted of a sequence of steps in the form of biochemical intermediates in which the output is connected to a downstream feedback loop. The gain in this network can have non monotonic behavior in which the low frequency components of the signal are also filtered at time scales commensurate with the induction of the regulatory loop. This behavior enables the network to filter out signals of long

duration. The minimal topological features of these biochemical networks provide distinct and robust mechanisms for integrating signals that persist with different characteristic time scales. As different temporally regulate signals often lead to different transcriptional programs such as in NF-κB signaling[93, 94], it is tempting to speculate on the role that such filtering mechanisms may have in regulating gene expression.

Chapter 4.4. Appendix

Chapter 4.4.1. Non monotonic frequency dependent signal integration by two competing and interacting products

In the main text in this chapter, a mechanism for integrating signals of long and short duration from different network topologies was proposed. Here, an alternative scheme for filtering of long duration signals is presented. This mechanism also illustrates the time scale dependence of signal transduction involving the competition of two interacting products. One product is produced in greater amounts at one time scale and the other is produced in great amounts at a different time scale. That is, a non monotonic, frequency dependent gain is obtained to filter low frequency or long duration signals. Consider the following scheme depicted in Fig. 4-5.

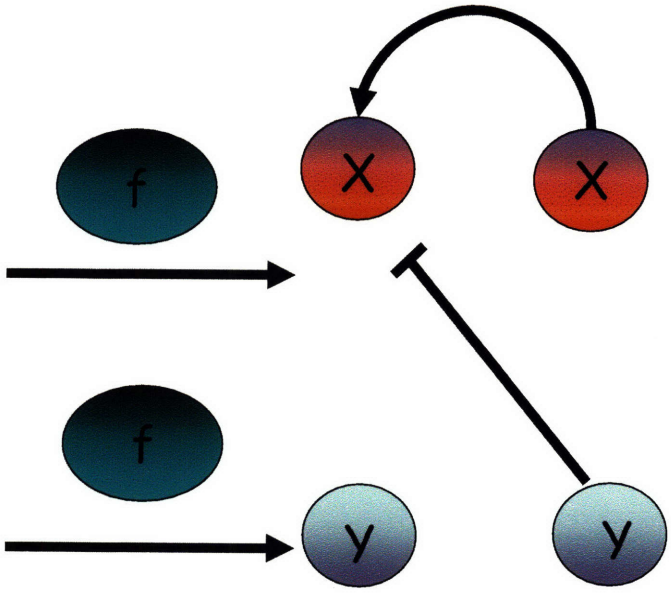


Figure 4-5. **A minimal model of two interacting gene products produced by the same signal**

Two species, denoted by X and Y in the cartoon respectively, are produced by the same signal f. Species Y negatively interacts with species X. Species X can positively interact with itself. The activity of species f is transient with associated signal duration.

We consider the dynamics of two interacting products that are produced by the same signal. A set of two kinetic equations can be written as follows:

$$\begin{aligned} \frac{dx}{dt} &= \alpha_1 f(t) - \beta_1 x - \gamma y;\ \beta_1 = \left(\beta_1^0 - k_f\right) \\ \frac{dy}{dt} &= \alpha_2 f(t) - \beta_2 y \end{aligned} \qquad \text{(4-A1)}$$

k_f is the strength of the positive feedback loop. γ is the strength of the negative interaction from Y to X. The positive feedback loop associated with species X can be modeled (in the linear regime) by decrease in the rate constant for degration β_1, i.e. $\beta_1 = \left(\beta_1^0 - k_f\right)$. β_1^0 is the rate constant for degradation of species X in the absence of

feedback. k_f sets the strength of the positive feedback. For the purposes of this study, differences in k_f are not considered. However, differences in k_f become important when considering the competition between positive and negative feedback interactions. Other parameters are also introduced in this model. α_1 and α_2 set the strength of interaction between the signal stimulus $f(t)$ and X and Y respectively. The bare degradation rates of X and Y respectively are β_1 and β_2.

We can solve equation 4.A1 using Fourier transformation as shown in detail in the main text. We define the following: $P_X(\omega) \equiv |X(\omega)|^2$ and $P_Y(\omega) \equiv |Y(\omega)|^2$; where,

$X(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} x(t)$ and $Y(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} y(t))$, and $S(\omega) = |F(\omega)|^2$ and $F(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} f(t)$:

$$P_X(\omega) = g_X(\omega) S(\omega) \qquad \text{(4-A2)}$$

$$P_Y(\omega) = g_Y(\omega) S(\omega) \qquad \text{(4-A3)}$$

After some algebraic manipulation, the frequency dependent gain for species X and Y are obtained,

$$g_X(\omega) = \frac{(\alpha_1)^2\left[(\beta_2)^2 + \omega^2\right] + \gamma\alpha_2(\gamma\alpha_2 - 2\alpha_1\beta_2)}{\omega^4 + \omega^2\left[(\beta_1)^2 + (\beta_2)^2\right] + (\beta_1\beta_2)^2} \qquad \text{(4-A4)}$$

$$g_Y(\omega) = \frac{(\alpha_2)^2}{\omega^2 + (\beta_2)^2} \qquad \text{(4-A5)}$$

Now consider plots of $g_X(\omega)$ (blue) and $g_Y(\omega)$ (green):

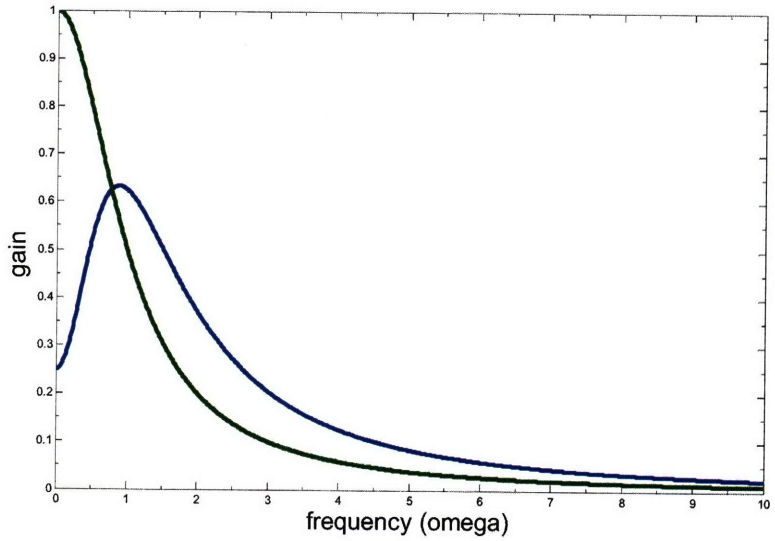


Figure 4-6. **Frequency dependent gain of two interacting products.**

X (blue, solid line) and Y (green, solid line). All parameters are taken to be 1.0 (in appropriate units) save α_2; $\alpha_2 = 1.5$.

From Fig. 4-6, the following behavior is apparent. At very high frequencies, signals coming from $f(t)$ are filtered by both products X and Y. At intermediate frequencies ($\omega \sim 1.0$ to $\omega \sim 10.0$) signals deriving from $f(t)$ are integrated more efficiently by species X. Therefore, short duration signals are integrated more efficiently by species X. At low frequencies, $\omega < 1.0$, signals originating from $f(t)$ are integrated more efficiently by species Y. Therefore, long duration signals are integrated more efficiently by species Y since at long times, X is affected by the negative interaction from Y.

This simple two species model illustrates how signal specificity can be achieved from two competing products by introducing changes in signal duration. Short duration signals are more effectively integrated by one species and long duration signals are more effectively detected by the other product.

Chapter 5 – Allovalency Revisited: An Analysis of Multisite Phosphorylation and Substrate Rebinding

Chapter 5.1. Introduction

The establishment of precise controls within signaling modules is an evolutionary prerequisite for a robustly functioning cellular system. A central issue to such control is the regulation of a dose response or the necessary input-output relationships that direct a specific biological function[10, 12, 16, 32, 95]. One such input that is widely utilized in many biological systems is the number of phosphorylations on a protein containing many potential phosphorylation sites. Multisite phosphorylation is ubiquitous in cell biology and regulates myriad cell decisions[96-102].

One salient example comes from the regulation of the cell cycle by ubiquitin mediated protein degradation, a key motif in the control of the cell cycle[102-108]. In the seminal work by Nash et al.[109] , the authors show that the CDK inhibitor, Sic1 functions through a thresholding mechanism – Sic1 must be phosphorylated at least 6 six (of its 9 possible) sites in order to be ubiquitinated and subsequently targeted for degradation. Sic1 is intrinsically disordered[110] and the location and specificity of these six phosphorylation sites seems to be unimportant at least to some extent. This observation among others[111] led to the hypothesis that the function of these seemingly redundant post translational modifications may be to increase the probability that Sic1 rebinds to its substrate upon disassociation[112, 113] and a mathematical model[112] was developed to investigate the rebinding of a polyvalent ligand. In this model, a ligand, once disassociated, effectively escapes from its receptor unless it is phosphorylated a sufficient number of times so as to increase its chances of rebinding.

The problem of ligand rebinding has been extensively studied in many contexts[114-120]. Some of the most comprehensive studies were carried out in the

context of two settings: 1.) ligand binding/unbinding to and from a surface as a model for the kinetics of ligand binding to cell-surface receptors[115, 117, 121-123] and 2.) chemotaxis and autocrine signaling resulting in rebinding of a ligand secreted from a cell[114, 120, 124, 125]. In each of these studies, it was demonstrated that ligand rebinding can be very significant. Despite these advances, how changes in the phosphorylation state of a substrate is related to rebinding and how this affects a biological dose response curve has not been fully investigated. A schematic of this effect is shown in Fig. 5-1.

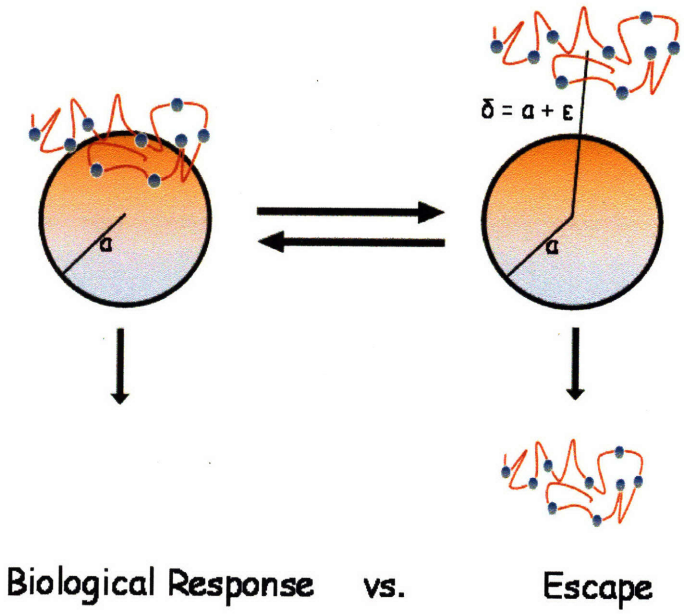


Figure 5-1. **How biological responses might be shaped by allovalency, multisite phosphorylation, and ligand rebinding.**

A schematic for a ligand, with multiple equivalent binding sites, potentially rebinding to its enzyme. Once the ligand unbinds from its target, two possible outcomes are available: 1.) escape from its binding partner (i.e. diffuse a distance far away from the receptor) and 2.) immediate rebinding to its receptor. A biological response can then be initiated if the protein is bound sufficiently long. The outcome is expected to depend on the number of sites (in the form of phosphorylations of the protein) that are available. Circles

depict different potential binding sites that arise from phosphorylations.

Towards this end, we use an integral equation theory and Monte Carlo simulations to study the rebinding of a ligand to a receptor, from which it initially disassociated, and how this rebinding may be affected by multiple recognition sites. From considering only the effects of a single molecule rebinding to its receptor, we compute the time dependence of the probability that a ligand remains bound as a function of the number of phosphorylations. In turn, we compute the probability that a ligand escapes its target as function of the number of recognition sites. The model and numerical simulations predict that this escape probability can decrease nearly exponentially as a function of the number of independent binding sites thus suggesting that ligand rebinding greatly affects the binding kinetics. We also highlight the importance of two physical regimes of ligand rebinding that are characterized by weak and strong rebinding and show how each regime may affect the input-output relationships of a system with multiple phosphorylation sites. We further note that the model predicts that, although a ligand's propensity to immediately rebind, as a function of the number of available binding sites, greatly affects the shape of the biological response, additional mechanistic ingredients appear to be required to achieve a highly cooperative response. Finally, we note that while our model predicts that the probability of a polyvalent ligand escaping from its receptor decreases exponentially as a function of the number of binding sites, this property appears insufficient to give rise to a highly cooperative response as has been previous predicted[112]. The source of this discrepancy appears to lie in how

the rate constants in the previous phenomenological model were varied independently to achieve the desired cooperativity.

Chapter 5.2. Methods and model development

Chapter 5.2.1. Multisite phosphorylation and ligand rebinding

The key considerations that are used to develop our model lie in the questions that we wish to address in this study. In particular, our aim is to investigate how ligand rebinding may be affected by multisite phosphorylation. Other studies of multisite phosphorylation have investigated the consequences of other physical effects such as distributive phosphorylation and feedback regulation[1, 101, 126]. We are interested in computing the probability that a ligand remains bound as a function of time and as a function of the number of recognition sites on the receptor.

To model this scenario, we assume that at time zero, a ligand is bound to its receptor and can be released with a constant unit time probability. When the ligand is in immediate proximity of the receptor, there is a probability θ that the ligand rebinds to the receptor within the time it takes to diffuse away from the immediate vicinity of the recognition domain. Multiple phosphorylations are then parameterized by a change in this probability. In the case we consider, which we refer to as the 'allovalent' model[112], each phosphorylation contributes equally and independently to the value of the parameter θ; i.e.

$$\theta = n\theta_0,$$

where n is the number of phosphorylations and θ_0 is the probability that a ligand that is proximally located to the recognition site will rebind when it is singly phosphorylated on

any site. In the work by Klein et al[112], this assumption (that each site contributes equally and independently to the rebinding probability) was sufficient to give rise to a highly cooperative response. Our aim is to further investigate the consequences of such an assumption.

Important to note is that in order for θ to be a probability it must be less than or equal to one. Therefore, θ_0 must be bounded by $\frac{1}{N}$;

$$\theta_0 \leq \frac{1}{N},$$

where N is the maximum number of phosphorylation sites on the ligand. An additional complication that is not considered here is the time dependence of n that may become important at late times. The theory therefore aims to investigate solely how rebinding is affected given a fixed number of binding sites. Also, this description of ligand binding is considered to be a "mean field" treatment since all conformational fluctuations of both the ligand and its receptor are neglected by the introduction of the parameter θ. One could also imagine that θ could have a complex, nonlinear dependence on n for a given θ_0 (i.e. $\theta = f(n;\theta_0)$) as would be the case when cooperative electrostatic interactions among the multiple phosphate groups influence binding[127].

Chapter 5.2.2. A self consistent integral equation theory for ligand rebinding

To begin our analysis, we exploit a formalism that monitors the trajectories of a single ligand as it disassociates from and potentially rebinds to its target to which it was originally bound. The formalism was developed by Tauber et al.[117] who investigated

the effects of a ligand binding to a collection of receptors on a planar surface in the context of surface plasmon resonance studies. Although our approach is similar in many regards, there are some subtleties that distinguish the two approaches and they are discussed within our treatment. We consider an equation that describes the time-evolution of the probability, $f(t)$, for a single ligand to be bound to its receptor provided that it is initially bound to its target. The initial condition $f(0)=1$ is used. Single molecule master equations of this sort have been used extensively in many different contexts[128, 129].

A knowledge of this function allows one to compute the probability that a ligand is bound as a function of time as well a time dependent escape probability which is taken to be, $1-f(t)$. A differential equation for the time evolution of $f(t)$ can be written as

$$\frac{df}{dt}=\upsilon_{+}-\upsilon_{-} \qquad (5\text{-}1)$$

The negative contribution υ_{-} simply follows first order disassociation kinetics (i.e. $\upsilon_{-}=k_{-}f(t)$). Thus, in the absence of rebinding ($\upsilon_{+}=0$), $f(t)$ decays via a single exponential with time constant $\frac{1}{k_{-}}$, $f(t)=e^{-k_{-}t}$. υ_{+} on the other hand is entirely due to the contribution from the rebinding of a single previously disassociated ligand. The forward rate of binding, υ_{+}, is therefore the probability that a protein dissociates in the interval τ and $\tau+d\tau$ and then subsequently rebinds at a later time interval, $t-\tau$ and $(t-\tau)+d\tau$, integrated over all previous times, τ. An equation for υ_{+}, therefore, can be written as follows:

$$\upsilon_{+} = k_{-}\int_{0}^{t} d\tau f(\tau) R(\delta, t-\tau) \qquad (5\text{-}2)$$

$R(\delta,t')$ is the probability per unit time that a protein binds to its target in the time interval $\{t',t'+dt\}$ given that it is located a distance, δ, away from the target at time 0 (δ is the small distance from the receptor that the ligand is placed when it disassociates).

Combining eqs. 1 and 2, we obtain an integral equation that accounts for the state of the ligand as a function of its entire history:

$$\frac{df(t)}{dt} = k_{-}\left[\left\{\int_{0}^{t} d\tau f(\tau) R(\delta, t-\tau)\right\} - f(t)\right]. \qquad (5\text{-}3)$$

We can analyze eq. 3 first by introducing Laplace-transformed variables:

$$\tilde{f}(s) = \int_{0}^{\infty} dt e^{-st} f(t)$$

and

$$\tilde{R}(\delta,s) = \int_{0}^{\infty} dt e^{-st} R(\delta,t).$$

By substituting the Laplace transforms into eq. 3 and making use of the convolution theorem[130], we obtain:

$$s\tilde{f}(s) - f(0) = k_{-}\tilde{f}(s)\left[R(\delta,s) - 1\right]. \qquad (5\text{-}4)$$

Or, upon rearranging and inverting the Laplace transform:

$$\tilde{f}(s) = \frac{f(0)}{s + k_{-}\left[1 - R(\delta,s)\right]}$$

$$f(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} ds e^{st} \frac{f(0)}{s + k_{-}\left[1 - R(\delta,s)\right]}. \qquad (5\text{-}5)$$

Thus, the probability that a protein remains bound can be solved exactly provided that an explicit form of $R(\delta,s)$ can be obtained and that the resulting Laplace inversion can be computed.

A convenient way to obtain $R(\delta,s)$ as developed previously[117] is to compute the quantity self-consistently by considering the statistics of first passage processes for an individual protein disassociating from its ligand: i.e,

$$R(\delta,t)=\theta F(\delta,t)+(1-\theta)\int_0^t d\tau R(\delta,t-\tau)F(\delta,\tau) \qquad (5\text{-}6)$$

where θ is a parameter that gives the probability that the protein will bind to its substrate given that it is within a distance δ; and, $F(\delta,t')dt'$ is the probability that a protein first reaches the origin, starting from a distance δ at time 0, in the time interval $\{t',t'+dt'\}$. In the case we study, θ is a linear function of the number of phosphorylations n, $\theta=n\theta_0$ where θ_0 is the probability that a ligand binds given that it has been singly phosphorylated. The contribution of the first term in eq. 5-6 is from the probability that a ligand is absorbed the first time it reaches its target. The contribution of the second term is from the probability that the ligand reached the target at time $\{\tau,\tau+dt\}$, was reflected (i.e. the ligand did not bind before it diffused away) at that time, and was then later absorbed at $\{t-\tau,(t-\tau)+dt\}$.

Again, upon Laplace transforming eq. 5-6 and the first passage time PDF, i.e. $\tilde{F}(s)=\int_0^\infty dte^{-st}F(\delta,t)$, and again, noting the convolution theorem, eq. 6 becomes:

$$\tilde{R}(\delta,s)=\theta\tilde{F}(\delta,s)+(1-\theta)\tilde{R}(\delta,s)\tilde{F}(\delta,s)$$

$$\tilde{R}(\delta,s)=\frac{\theta\tilde{F}(\delta,s)}{1-(1-\theta)\tilde{F}(\delta,s)}. \qquad (5\text{-}7)$$

In the work by Tauber and coworkers[117], a similar equation as eq. 5-7 is used to compute $R(\delta,t)$. However, in their treatment of the calculation of $R(\delta,t)$, the coefficients θ and $1-\theta$ in eq. 6 are replaced with $\theta(1-f(t))$ and $1-\theta(1-f(t))$. In their problem, the authors considered rebinding to receptors on a planar surface and the probability that a ligand reaches a receptor that contains a ligand that already contains a bound ligand need be taken into account. Our equation for the absorption probability does not require the additional $1-f(t)$ factor since we are only considering the rebinding of a receptor to a single isolated receptor.

For further analysis, the first-passage time distribution function is now required and is considered in three dimensions. Assuming spherical symmetry, the solution to the first passage problem can be obtained in the Laplace domain and its derivation is contained in the appendix; thus,

$$\tilde{F}(a+\varepsilon;s)\approx\left[\frac{a}{a+\varepsilon}e^{-\sqrt{\tau s}}\right]. \qquad (5\text{-}8)$$

The distance, δ, is written as $\delta=a+\varepsilon$ (a is the radius of the recognition domain and ε is the average distance away from the boundary of the recognition domain that the ligand is initially displaced when it rebinds) the variable $\tau=\frac{(a+\varepsilon)^2}{D}$ has been introduced along with D being the diffusion constant of the ligand. τ is the time it takes for the ligand to diffuse a distance on the order of the distance to its target.

A further simplification can be made if we observe the system on time scales commensurate with the disassociation time; $t \sim 1/k_{-}$, i.e. $t >> \tau$ (so that s is small).

$\tilde{F}(a+\varepsilon;s)$ becomes:

$$\tilde{F}(a+\varepsilon,s) \approx \frac{a}{a+\varepsilon}\left[1-\sqrt{\tau s}\right]+O(\tau s). \tag{5-9}$$

This approximation has been shown to be very good in one dimension[117] in which rebinding is believed to be more prominent. Therefore, up to order $O(\tau s)$, we substitute eq. 5-9 into eq. 5-7 and obtain:

$$1-\tilde{R}(a+\varepsilon,s) \approx \frac{(1-\gamma)+\gamma\sqrt{\tau s}}{1-(1-\theta)\gamma} \tag{5-10}$$

where $\gamma = \frac{a}{a+\varepsilon}$.

Inserting this expression into the integrand in eq. 5-5 yields:

$$\begin{aligned} \tilde{f}(s) &= \frac{f(0)}{s+k^{eff}_{-}\left[(1-\gamma)+\gamma\sqrt{\tau s}\right]} \\ f(t) &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} ds \frac{e^{st} f(0)}{s+k^{eff}_{-}\left[(1-\gamma)+\gamma\sqrt{\tau s}\right]} \end{aligned} \tag{5-11}$$

where, $k^{eff}_{-} = \frac{k_{-}}{1-(1-\theta)\gamma}$.

Chapter 5.2.3. Monte Carlo simulations

To supplement the theory, we also considered Monte Carlo simulations. Simulations were performed by considering a collection of random-walkers with a set of receptors on a three-dimensional lattice of 100 x 100 x 100 lattice spacings. Each protein

(receptor and ligand) occupy one site on the lattice at any given time. In each Monte Carlo step, with equal probability for a move to be made in any direction, an attempt to allow a molecule to diffuse is given by P_{diff} which defines a time sale that then defines a diffusion constant; i.e. $P_{diff} \sim D/L^2$ where D is the diffusion constant and L is the length of a lattice spacing which is taken to be the diameter of a typical protein or in this case, $L \square 10nm$. When encountering an immobile receptor at any of its nearest-neighbor positions, the substrate can bind with probability $P = P_{rxn} e^{-\left(\frac{E_{k_+}}{k_b T}\right)}$, so that $k_+ \sim e^{-\left(\frac{E_{k_+}}{k_b T}\right)}$. $k_b T$ is Boltzman's thermal energy, E_{k_+} is the energy barrier for association when a receptor and ligand come into contact. In this scheme the rebinding probability θ behaves as, $\theta = n\theta_0 \sim e^{-\left(\frac{E_{k_+}}{k_b T}\right)}$.

The fraction of bound ligands was computed by sampling at steady-state, as a function of θ, $\theta \propto k_+$. Escape probabilities were computed by first allowing a receptor to release its ligand at time $t = 0$; at a later time, $t = t_0$, sampling of whether or not the ligand is again bound to its target is performed. t_0 was chosen to be a time on the order of the encounter time for a protein in a eukaryotic cell; $t_0 = 1000 mcsteps$ ($1000 mcsteps \sim 1ms$ assuming a lattice spacing of $L = 10nm$ and a diffusion constant $D = 10\mu m^2/s$). For each value of θ, the statistics determining the escape probability were obtained from 100,000 independent trials.

Chapter 5.3. Results and discussion

Chapter 5.3.1. Rebinding probabilities

From eq. 11, the relevant biological quantities can be computed. First consider the absorption probability in the Laplace domain. A numerical inversion of eq. 11 can in principle be accomplished and the subsequent function plotted. However, since such a computation is difficult to accomplish due to numerical instabilities resulting from the multi-scale nature of the computation, we considered the function in the Laplace domain. By substituting the results contained in eq. 8 into the expression for $R(\delta,s)$ (eq. 5-7), we obtain.

$$\tilde{R}(\delta,s)=\frac{\theta\gamma e^{-\sqrt{\tau s}}}{1-(1-\theta)\gamma e^{-\sqrt{\tau s}}} \tag{5-12}$$

As seen in Fig. 5-2a, since the first-passage time distribution decays as a stretched exponential function in the Laplace domain, rebinding can be significant over many time scales.

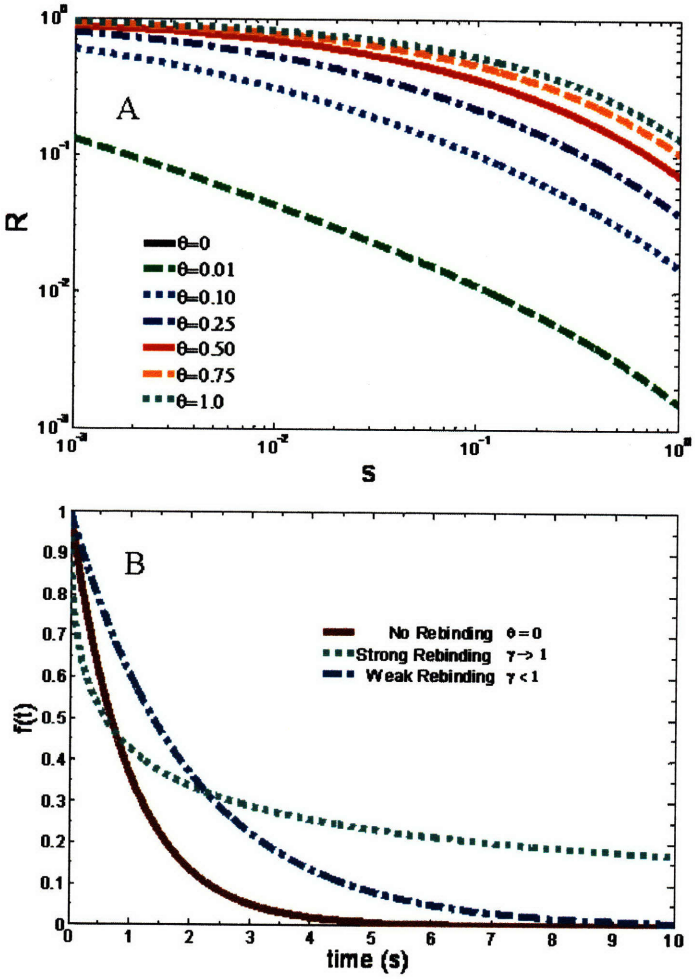


Figure 5-2. **Strong ligand rebinding can be significant over many time scales.**

a.) Plots of the absorption probability in the Laplace domain, $R(\delta;s)$, with units chosen so that the microscopic diffusion time scale τ is unity, $\tau = 1$; ($\tau = \frac{(a+\varepsilon)^2}{D}$), are shown on a log-log plot. The strong rebinding ($\gamma \to 1$) limit is considered for convenience. $R(\delta;t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} ds R(\delta;s)e^{st}$ is the probability that a ligand absorbs to its target a distance δ away at time t. Plots are generated from the expression obtained using eq. B2 . b.) Shapes of the dissociation curves in three limits: 1.) when no binding occurs, 2.) when $\gamma \to 1$ (strong rebinding), and 3.) when $0 < \gamma < 1$ or

$\varepsilon = O(a+\varepsilon)$ (weak rebinding). Dashed lines show the behavior of the decay curve in the absence of rebinding, $k_- = 1$. Dotted lines give the case when the decay curve for rebinding takes the form of the strongly non-exponential one-dimensional case i.e. $\gamma = 1$. The time constant, κ ($\kappa = \frac{4k_-^2\alpha}{\theta^2}$), in the appendix is taken to be unity $\kappa = 1$. Dash-dotted lines show the behavior of the decay curve in the instance of weak rebinding limit ($k_-^{eff}(1-\gamma) = \frac{1}{2}$).

Chapter 5.3.2. Kinetics of disassociation modified by rebinding events—exponential versus non-exponential decay giving rise to 'strong' and 'weak' regimes of rebinding

In one dimension, for all parameter ranges, rebinding events lead to strongly non-exponential kinetics whenever significant rebinding is possible (Appendix). That is, as a result of rebinding, a ligand can remain bound to its receptor long after the time scale that characterizes its dissociation. In three dimensions, the effects of rebinding should be less significant since fewer returns to the origin occur in higher dimensions and some trajectories never return to the origin[75].

Upon inspection of the Laplace inversion in eq. 5-12, two kinetic regimes are observed that depend on the relative size of the receptor as determined by $\gamma = \frac{a}{a+\varepsilon}$. First, if

$$(1-\gamma) << \gamma\sqrt{\tau s} \qquad (5\text{-}13)$$

(e.g. the radius of gyration of the disordered protein is small compared to the radius of the region to which it binds to its targeted substrate, $\varepsilon \approx 0$ and $\gamma \sim 1$), then the overall kinetics of ligand disassociation that are modified as a result of rebinding events behave in a similar fashion to that of the one-dimensional case[117] as shown in the appendix. This can be seen by taking the $\varepsilon \to 0$ (i.e. $\gamma \to 1$) limit of eq. 11 in which case,

$$f(t) \to f(0) e^{\kappa t} erfc\left(\sqrt{\kappa t}\right) \qquad (5\text{-}14)$$

where, κ^{-1} is a time scale that behaves as $\kappa \to \frac{4k_-^2 \alpha}{\theta^2}$ as $\gamma \to 1$ (appendix).

One the other hand, for γ significantly less than one, $(1-\gamma) >> \gamma\sqrt{\tau s}$, since τ is a microscopic time scale, an exponential decay is observed:

$$f(t) \square e^{-k_d t} \qquad (5\text{-}15)$$

where $k_d = \frac{k_-(1-\gamma)}{\left[1-\gamma(1-\theta)\right]}$.

In Fig. 5-2b, plots of the decay of the probability $f(t)$ are shown for three cases. In the first case, no rebinding binding ($\theta = 0$) is considered and $f(t)$ behaves according to: $f(t) = f(0)e^{-k_- t}$. In the second case, strong rebinding is considered ($\gamma \to 1$) so that $f(t)$ takes on highly non-exponential behavior; i.e. $f(t) = f(0)e^{\kappa t} erfc\left(\sqrt{\kappa t}\right)$. Finally, in the third case, weak rebinding is considered ($\gamma < 1$) so that $f(t)$ takes the form: $f(t) = f(0)e^{-k_d t}$. The parameter values used are given in the figure caption. As shown in the plot, the two regimes of rebinding lead to dramatically different consequences. When

the ligand begins significantly far away from its target leading to the weak ($\gamma < 1$) rebinding regime, rebinding serves simply to decrease the off rate ($k_d < k_-$). In contrast, when the ligand begins close to its target (or the target is very large in comparison to the ligand), $\gamma \to 1$ and the shape of the disassociation curve changes dramatically resulting in nonexponential disassociation kinetics. The presence of a distribution with a fat tail (i.e. $\sim t^{-1/2}$; $t >> (1/\kappa)$) is observed; this signifies that the release of the ligand is distributed over many time scales – the ligand becomes trapped by the receptor for long times.

Chapter 5.3.3 The fraction of bound ligands can be greatly influenced by rebinding

With the formulas obtained in eqs. 5-14 and 5-15, $f(\theta, t_0)$, the probability that a ligand remains bound as a function of θ, can be studied at different time points, t_0. Shown in Figs. 5-3a, b, the behavior of these functions is plotted. For the strong rebinding ($\gamma \to 1$) case in Fig. 5-3a, it can be seen that the fraction of bound ligands is strongly influenced by rebinding over a broad range of time scales (i.e. 0.001s – 1000s). On the other hand, for weak rebinding, the fraction of bound ligands is only strongly influenced by rebinding on a time scale, τ_- commensurate with the intrinsic off-rate (i.e. $\tau_- \approx \frac{1}{k_-}$). Such behavior is a direct consequence of the non-exponential vs. exponential shapes of the decay curves. It is also noted that fitting each curve to a Hill function $\frac{\theta^H}{K_{50\%}^H + \theta^H}$ by nonlinear

regression, gives a value of $H \approx 1$ for all curves indicating a 'Michaelian' dose response[32].

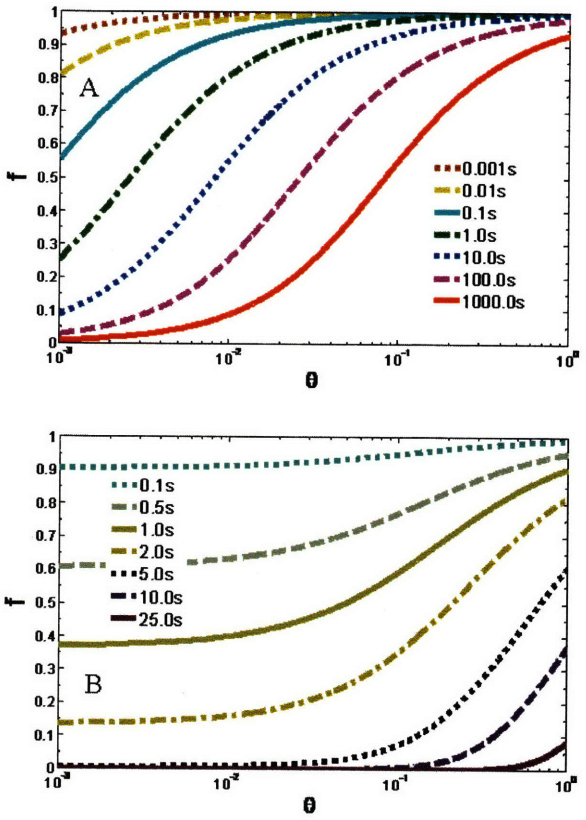


Figure 5-3. **Rebinding is influenced by an increase in the number of phosphorylation sites.**

$f(\theta,t_0)$ is plotted for different values of t_0 given on the legend: the two regimes a.) strong rebinding, $\gamma \to 1$ and b.) weak rebinding regime, e.g. $\gamma = 0.9$; for both instances, $\tau = 10^{-6} s$, $k_- = 1 s^{-1}$. $f(\theta,t_0)$ gives the probability that a ligand remains bound to its target as a function of the number of phosphorylation sites, θ, and at a given time t_0. When the time t_0 is commensurate with or greater than the intrinsic time constant $\frac{1}{k_-}$, i.e. $t_0 > \frac{1}{k_-}$, the positive contribution to the function, $f(\theta,t_0)$ is mostly due to rebinding.

Chapter 5.3.4. Escape probabilities and the effects of rebinding on dose response curves

The escape probability can be computed within this theory from a consideration of the fraction of bound ligands $f(\theta,t_0)$. $1-f(\theta,t_0)$ gives the probability that a ligand is not bound to its target at time t_0 (i.e. the probability that the ligand has "escaped"). As seen in the plots in Figs. 3a and 3b, for large enough values of θ, long after the disassociation from the first order decay process, ligands can be trapped by their receptors.

In the weak rebinding regime, the escape probability has the functional form: $1-f(\theta,t_0)\approx 1-\exp\left(-\frac{a}{[b+c\theta]}t_0\right)$ as can be seen upon rearranging eq. 5-15. On the other hand, in the strong rebinding regime (eq. 5-14), the escape probability behaves as: $1-f(\theta,t_0)\approx 1-e^{at_0\theta^{-2}}erfc\left(a\theta\sqrt{t_0}\right)$. For typical parameter values, these functions decay at rates commensurate with the rates of an exponential process characterized by a single time scale as seen in Figs. 3a and 3b.

Alternatively, Monte Carlo simulations[35] can be used to compute the escape probabilities numerically. Plots of the escape probabilities are shown in Fig. 4a; as indicated on the inset of the plot, the data obtained from the Monte Carlo simulations are shown to fit well to an exponential decay function with a single parameter i.e. $f(\theta,t_0)\square e^{-k\theta}$. The fraction of receptors bound as a function of θ is also computed from the computer simulations and plotted in Fig. 5-4b. Different values of receptor density are considered. For each curve, as exemplified on in the inset of Fig. 5-4b, a fit to

a Hill function gives a Hill coefficient of near unity. The plots in Fig. 5-4b. are consistent with those obtained from the theory and plotted in Fig. 5-3a.

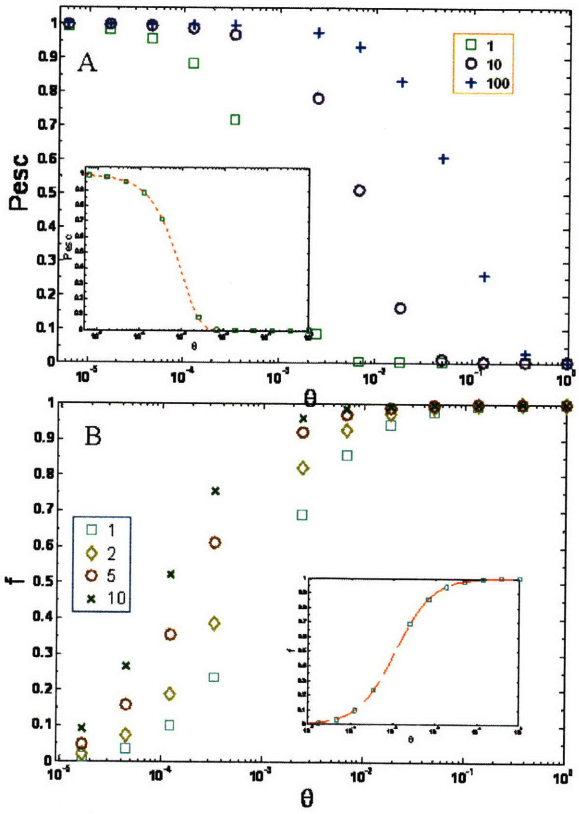


Figure 5-4. **Monte Carlo simulations suggest that an exponential decrease in the escape probability for an increasing number of phosphorylation sites can be insufficient to produce a switch-like dose response.**

Plots of simulation data from Monte Carlo simulations are shown. a.) The escape probability, *Pesc* (defined in the methods section), as a function of θ ($\theta \sim e^{-\left(\frac{E_{k_+}}{k_b T}\right)}$, $\theta = n\theta_0$) is given. Three different values of the effective diffusion constant $D_{eff} \equiv \frac{P_{diff}}{P_{rxn}}$ are shown: $D_{eff} = 1$ (squares), $D_{eff} = 10$ (circles), and $D_{eff} = 100$ (crosses). The plot in the inset contains a fit to an exponential function $Pesc = e^{-k_{esc}\theta}$ for the $D_{eff} = 1$ case; $k_{esc} = 10^3 mcsteps^{-1}$ was used in the plot. b.) The fraction of bound ligand as a function of θ is

shown. Four values of a scaled receptor density $\frac{\rho}{\rho_0}$, where $\rho_0 = 1000 receptors/cell$, are considered: $\rho = 1$ (squares), $\rho = 2$ (diamonds) $\rho = 5$ (circles) $\rho = 10$ (crosses). The plot in the inset gives a fit to a Hill function, $\frac{\theta^H}{K_{50\%}^H + \theta^H}$ with H = 1, for the case of $\rho = 1$. Error bars from the simulations are on the order of 5% of the reported values.

While the curves in Figs. 5-3a,b and 5-4b show that the Hill coefficient is near unity, thresholding effects in the dose response curves may appear when rebinding is significant. This thresholding effect that is observed in Figs. 5-3a, b, and 5-4b is defined as a different value of θ needed to reach a given value of $f(\theta, t_0)$. These results are thus similar to the observations that have been previously reported[101] that considered the case of multiple phosphorylation steps that occur in an ordered, distributive manner. This result is therefore expected to become more prominent upon incorporation of the possibility of rebinding.

Finally, we considered how the fraction or probability that a ligand remains bound vary as a function for the number of phosphorylations, n for different values of θ_0 (recall: $\theta = n\theta_0$). Four cases are shown: the strong rebinding ($\gamma \to 1$) case at long ($100s$) and short ($10s$) times (Figs. 5-5a,b), and the weak rebinding ($\gamma < 1$) at long ($5s$)and short ($1s$) times (Figs. 5-5c,d). As seen, graded responses are observed in each of these cases. Perhaps interesting to note is the non-uniformity of these dose response curves; some appear near linear while others have a nonlinear, hyperbolic shape. This effects results from a rescaling of θ_0. For different values of θ_0, the response to changing values of n is different.

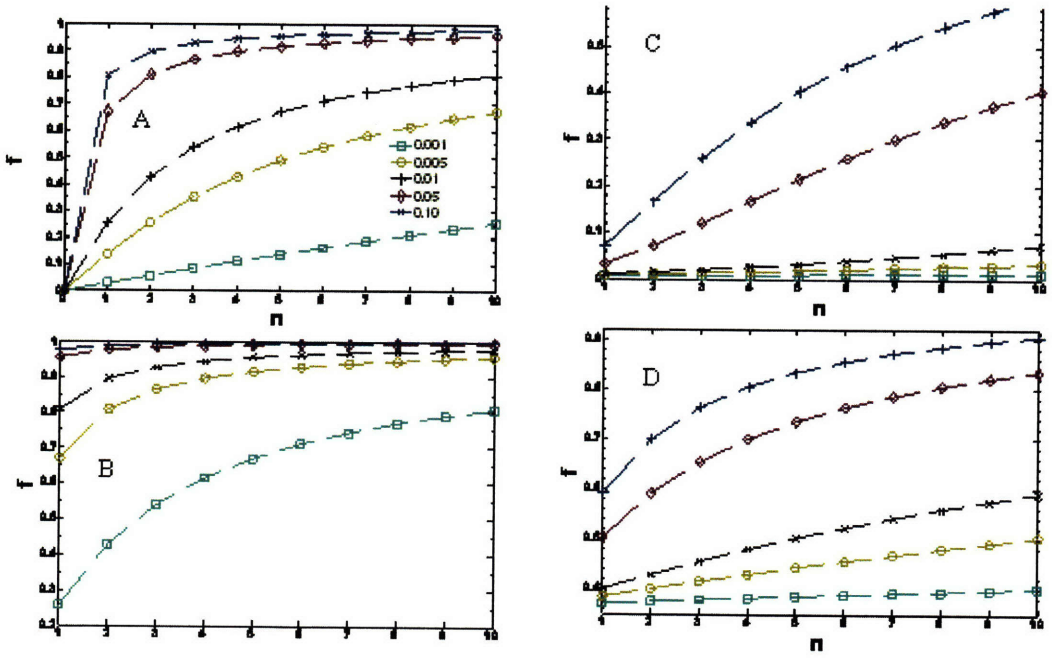


Figure 5-5. **Graded responses are observed for over wide ranges of parameter values.**

Plots of $f(n,t_0)$ in which $\theta = n\theta_0$ are shown for different values of θ_0. The number of phosphorylations, n is plotted along the abscissa. Strong (a,b) and weak (c,d) rebinding limits are considered. Numbers on the legend indicate the different values of θ_0 that were used. In the strong rebinding cases (a,b), two time points, t_0, are given: a.) $t_0 = 100s$ and b.) $t_0 = 1s$. In the weak rebinding cases (c,d), the two values of t_0 used were: c.) $t_0 = 5s$ and d.) $t_0 = 1s$.

Chapter 5.3.5. Comparison to previous theoretical work on ligand rebinding and multisite phosphorylation

Previous theoretical work has also studied the effects of multisite phosphorylation. In a prior study[112], a theoretical model predicted an exponential decay in the escape probability as a function of the number of phosphorylations. This exponential decay was predicted to be sufficient to give a highly cooperative dose

response curve (the addition of a single independent binding site results in a large increase in the fraction of ligands bound to their receptors). The model that was developed consists of a ligand existing in one of three states: bound to the receptor (B), in a region proximal to the receptor (P), and a region far away from the receptor (F). Transitions between these states are considered that result in the following kinetic scheme with kinetic constants, $k_{on}, k_{off}, k_{esc}, k_{cap}$;

$$B \underset{k_{on}}{\overset{k_{off}}{\rightleftharpoons}} P \underset{k_{cap}}{\overset{k_{esc}}{\rightleftharpoons}} F.$$

The fraction of bound ligands is taken to be the proportion of ligands in the bound and proximal states (i.e. $f_B = \frac{B+P}{L_{tot}}$ where B is the number of ligands in the bound state, P is the number of ligands in the proximal state, and L_{tot} is the total number of ligands.). Using mass action kinetics, an expression for f_B as a function of the kinetic constants and number of available receptors R_f can be computed[112]:

$$f_B = \frac{1}{1+\kappa} \tag{5-16}$$

where,

$$\kappa = \frac{k_{esc}}{k_{cap}}\left[R_f\left(1+\frac{k_{on}}{k_{off}}\right)\right]^{-1}. \tag{5-17}$$

k_{off} is an off rate for dissociation of a bound ligand and k_{on} is an on rate that is taken to be proportional to the number of phosphorylations ($k_{on} \propto n$). k_{esc} is computed from first passage time statistics and is shown to decay exponentially $k_{esc} \propto e^{-an}$ as a function of the number of phosphorylations n. It is also assumed that k_{cap} is a diffusion limited rate

constant[119] and is taken to be independent of n [112]; that is, $k_{cap} = 4\pi R_0 (D_l + D_r) N_A$ where R_0 is the radius of the receptor, D_l and D_r are diffusion constants for the ligand and receptor respectively, and N_A is Avogadro's number.

From the expression, it is clear that, for some parameter values, a highly cooperative response[32] can obtained when k_{esc} decays exponentially for increasing n while keeping k_{cap} fixed. However, in the framework of the model, it is not clear why k_{cap} (a rate constant of diffusion limited capture for ligands to enter the proximal region near the receptor) is independent of n or k_{esc}. If more ligands are immediately rebinding to their receptors and as a result k_{esc} decreases, then fewer ligands are available to diffuse into the proximal region denoted by the $F \rightarrow P$ transition. k_{cap}, it seems, should decrease accordingly.

Consistent with this model, the theory and Monte Carlo simulations both predict a fast (and nearly exponential) decrease in the probability of escape for a ligand as a function of the number of phosphorylations as seen in Fig. 5-4a. This exponential decrease in the probability of escape of a newly disassociated ligand, however, appears insufficient to produce a highly cooperative response due to rebinding (Fig. 5-4b). The discrepancy between these two findings appears to lie in the assumption of a constant value of k_{cap} that was used. The Monte Carlo simulations show that the allovalent model predicts a Michaelian[32] (i.e. Hill coefficient of unity) response. This result is perhaps not surprising since there is no cooperativity introduced into the model. The effect of rebinding, in itself, appears insufficient to give a cooperative response. However, despite this apparent lack of cooperativity, differential rebinding effects (with respect to changes

in the number of binding sites) can be very significant as has been emphasized throughout this work.

Chapter 5.4. Summary

We first reformulated the problem of the rebinding of a protein with multiple independent phosphorylation sites, to its target in the context of a self consistent integral equation theory[116, 117], to study the effects of one dimensional ligand rebinding to a surface containing antibody receptors. Within this formalism, we solved the rebinding problem of a single ligand to an isolated receptor in three dimensions in two limits that depend on the relative sizes of the receptor and ligand. We find two qualitatively distinct regimes of rebinding kinetics whose crossover depends mainly on the size of the substrate and its target. In one regime (i.e. when there is strong rebinding), the kinetics of ligand disassociation takes on a similar functional form to that of the one-dimensional case—this results resulting in a slow decay of bound substrates characterized by non-exponential kinetics and a power law tail. Alternatively, in the other regime (i.e. when there is weak rebinding), the behavior of the kinetics of disassociation exhibits an exponential form and is thus characterized by a single rate constant – rebinding gives simply a slower time constant signifying a lesser influence on rebinding. The model predicts that the relative size of the ligand (that determines the rebinding regime) may play a key role in determining the functional role of multisite phosphorylations. It may be interesting to study how the different regimes of rebinding, that are predicted in this model, relate to other biological processes that require ligand rebinding at different length and time scales, such as autocrine signaling[131-133].

We then used the results obtained to compute rebinding probabilities. We showed that, in some instances, rebinding can occur over many time scales and contribute significantly to the total bound fraction of ligands. Furthermore within this model, an increase in the number of independently acting phosphorylation sites leads to a near exponential decrease in the probability that a ligand escapes from its target (i.e. it diffuses a large distance without being captured by its target). The model also predicts a graded response[32] and yields a Hill coefficient of near unity for all parameter values. Thus, statistically independent contributions to the association rate of the ligand in the form of additional binding sites and their additive effect on the association rate (while potentially having a great impact on the binding kinetics) does not appear to in itself yield a highly cooperative response. These additional binding sites can, however, influence the shape of the dose response in a nonlinear manner.

Previous theoretical work[112] has also studied the effects of multisite phosphorylation on substrate rebinding. This model also predicts an exponential decay in the escape probability as a function of the number of phosphorylations. This effect then gives rise to a highly cooperative dose response curve (the addition of a single independent binding site can result in a large increase in the fraction of ligands bound to their receptors). However, a high degree of cooperativity is not observed in both our theoretical treatment and Monte Carlo simulations.

Although rebinding may not, in itself, produce a 'switch-like' (i.e. highly cooperative) dose response curve[32] in the fraction of ligands bound, it is

nevertheless interesting to speculate on the ways in which the rebinding of a substrate to its receptor may affect myriad cellular processes. For instance, by controlling the probability of rebinding in the form of changing the number of phosphorylations on an enzyme, the degree of processive vs. distributive enzymatic modifications[32, 48] that comprise a multi-step pathway could be controlled. It is also possible that the parameter γ in our model that is determined by the relative size of the receptor and ligand and other structural features of the protein-protein interaction would be a key determinant in the number of processive versus distributive phosphorylation events[48].

Many mechanisms have been proposed (and some tested) that can account for switch-like dose responses involving proteins with multisite phosphorylations[117, 127, 134, 135]. In the language of our model, such effects would result in θ having a complex, nonlinear relationship with n and θ_0. It may be interesting to explore how these mechanisms containing phenomena such as decoy phosphorylation, entropically driven binding, or electrostatics may couple to the effects of ligand rebinding as studied here.

Finally, the explicit geometry of the binding sites was not considered in this work. Other theoretical works[124, 136] have shown that these effects can be important in polyvalent ligand binding. In future work, it may be interesting to investigate these geometrical aspects of multisite phosphorylation and ligand rebinding. Such a study might be accomplished, for example, by borrowing ideas from polymer physics[137], and considering the dynamics of a flexible polyvalent chain and its interaction with a substrate.

Chapter 5.5. Appendix

Chapter 5.5.1. First passage time statistics and rebinding in three dimensions

The rebinding problem is now considered in three dimensions. Assuming spherical symmetry, the solution to the first-passage problem can be obtained in terms of modified Bessel functions. We introduce the survival probability

$$\Phi(\eta;t) = \int_t^\infty dt' F(\eta,t') = 1 - \int_0^t dt' F(\eta,t') \qquad \text{(5-A1)}$$

so that,

$$F(\eta;t) = -\frac{d\Phi(\eta,t)}{dt}. \qquad \text{(5-A2)}$$

In the Laplace domain:

$$\tilde{\Phi}(\eta;s) = \int_0^\infty dt e^{-st} \Phi(\eta;t), \qquad \text{(5-A3)}$$

The first passage time PDF can be written as follows:

$$\tilde{F}(\eta;s) = \Phi(\eta,0) - s\tilde{\Phi}(\eta,s) = 1 - s\tilde{\Phi}(\eta,s) \qquad \text{(5-A4)}$$

where, $\Phi(\eta,0) = 1$ (the survival probability at time zero is defined as 1). The survival probability can be obtained by solving a backwards Kolmogorov equation[75, 138] that has the form of a diffusion equation

$$\frac{\partial\Phi(\delta,t)}{\partial t} = D\nabla^2\Phi(\delta,t),$$

or, in the Laplace domain:

$$D\nabla^2\tilde{\Phi}(\delta,s)=s\tilde{\Phi}(\delta,s)-1 \quad , \tag{5-A5}$$

(from hereon, length is scaled with respect to a diffusion length scale;

$\eta=\delta\sqrt{\frac{s}{D}}$; $\delta=a+\varepsilon$) with absorbing boundary condition,

$$\Phi(\eta_a,s)=0 \tag{5-A6}$$

where $\eta_a=a\sqrt{\frac{s}{D}}$ and a is the radius of the sphere containing the targeted substrate.

For the other boundary condition, far away from the target at a distance, η_0, $\Phi(\eta;t)$ is unity; i.e.

$$\tilde{\Phi}(\eta\to\eta_0;s)=\frac{1}{s}. \tag{5-A7}$$

In spherical coordinates, eq. A5 becomes:

$$\frac{d^2\tilde{\Phi}(\eta;s)}{d\eta^2}+\frac{2}{\eta}\frac{d\tilde{\Phi}(\eta;s)}{d\eta}-\tilde{\Phi}(\eta;s)+\frac{1}{s}=0 \tag{5-A8}$$

and has the general solution:

$$\begin{aligned}\tilde{\Phi}(\eta;s)&=\frac{1}{s}+A\frac{I_{-1/2}(\eta)}{\eta^{1/2}}+B\frac{I_{1/2}(\eta)}{\eta^{1/2}}\\ &=\frac{1}{s}+\sqrt{\frac{2}{\pi}}\left[A\frac{cosh(\eta)}{\eta}+B\frac{sinh(\eta)}{\eta}\right].\end{aligned} \tag{5-A9}$$

where $I_\nu(x)$ is a modified Bessel function of order ν.

The solution for $\tilde{\Phi}(\eta;s)$ that satisfies the boundary conditions in eqs. 5-A6 and 5-A7 gives the coefficients A and B:

$$A = \sqrt{\frac{\pi}{2}}\left(\frac{\eta_a}{s}\right)\frac{sinh(\eta_0)}{sinh(\eta_a - \eta_0)}$$

and

$$B = -\sqrt{\frac{\pi}{2}}\left(\frac{\eta_a}{s}\right)\frac{cosh(\eta_0)}{sinh(\eta_a - \eta_0)}.$$

Substituting the coefficients into eq. 5-A9 and making use of the appropriate trigonometric identities gives:

$$\tilde{\Phi}(\eta;s) = \frac{1}{s}\left[1 - \frac{\eta_a sinh(\eta_0 - \eta)}{\eta sinh(\eta_0 - \eta_a)}\right] \qquad \text{(5-A10)}$$

Now we assume that the length of the total system (i.e. the cell) is much larger than the length of a single protein ($\eta_0 >> \eta_a$); so that $sinh(\eta_0 - \eta_a) \approx sinh(\eta_0)$ and $tanh(\eta_0 - \eta_a) \approx 1$. Upon substituting these relations and performing some algebraic manipulations, we obtain:

$$\tilde{\Phi}(\eta;s) \approx \frac{1}{s}\left[1 - \frac{\eta_a}{\eta}\{sinh(\eta) - cosh(\eta)\}\right]$$

$$\approx \frac{1}{s}\left[1 + \frac{\eta_a e^{-\eta}}{\eta}\right] \qquad \text{(5-A11)}$$

Substituting eq. A11 into eq. A4 gives an expression for the first passage time, $\tilde{F}(\eta;s)$:

$$\tilde{F}(a+\varepsilon;s) \approx \left[\frac{a}{a+\varepsilon}e^{-\sqrt{\tau s}}\right] \qquad \text{(5-A12)}$$

where the distance δ is written as $\delta = a + \varepsilon$ and the variable $\tau = \frac{(a+\varepsilon)^2}{D}$ has been introduced. As in a prior study[117], a further simplification can be made if we observe

the system on time scales commensurate with signaling times (times over which signals are propagated); $t \sim \frac{1}{k_-}$, i.e. $t >> \tau$ (so that s is small); then $\tilde{F}(a+\varepsilon;s)$ becomes

$$\tilde{F}(a+\varepsilon,s) \approx \frac{a}{a+\varepsilon}\left[1-\sqrt{\tau s}\right]+O(\tau s). \qquad \text{(5-A13)}$$

Therefore, up to order $O(\tau s)$, we substitute eq. A13 into eq. 7 and obtain:

$$1-\tilde{R}(a+\varepsilon,s) \approx \frac{(1-\gamma)+\gamma\sqrt{\tau s}}{1-(1-\theta)\gamma} \qquad \text{(5-A14)}$$

where $\gamma = \frac{a}{a+\varepsilon}$.

Chapter 5.5.2. First passage time statistics and rebinding in one dimension

Although eq. 9 in 1d is exact, $\tilde{F}(\delta,s)$, however, often has a complicated form. Such a complication can make the Laplace inversion very difficult. For instance in the continuum limit in one dimension[75]:

$$F(\delta;t) = \frac{\delta}{(4\pi Dt)^{1/2}} \frac{e^{-\delta^2/4Dt}}{t} \qquad \text{(5-B1)}$$

which has the Laplace transform: $F(\delta;s) = e^{-\sqrt{\alpha s}}$ -- α is the microscopic time scale that it takes a protein with diffusion constant D to diffuse a tiny amount, δ; $\alpha = \frac{\delta^2}{4D}$.

Subsequently, eq. 11 can be substituted into eq. 9 to obtain:

$$\tilde{R}(\delta,s) = \frac{\theta e^{-2\sqrt{\alpha s}}}{1-(1-\theta)e^{-2\sqrt{\alpha s}}} \qquad \text{(5-B2)}$$

Despite this complication, additional simplifications can be made if we consider an observable time scale of signal transduction, $\tau_{sig} \sim (1/k\_)$, that is much longer than the microscopic diffusion time ($\alpha << \tau_{sig}$). In this case: $F(\delta;s) = e^{-\sqrt{\alpha s}} \approx 1 - \sqrt{\alpha s} + O(\alpha s)$. So that upon substituting into eq. B2, we obtain:

$$R(\delta;s) \approx 1 - \frac{2\sqrt{\alpha s}}{\theta}. \qquad \text{(5-B3)}$$

As in a previous study[117], substituting eq. 5-B3 into eq. 5-5 gives:

$$\tilde{f}(s) = \frac{f(0)}{s + (2k\_\theta)\sqrt{\delta s}}. \qquad \text{(5-B4)}$$

eq. B4 can be inverted[139]:

$$f(t) = f(0) e^{\kappa t} erfc\left(\sqrt{\kappa t}\right) \qquad \text{(5-B5)}$$

where $1/\kappa$ is a single characteristic time-scale ($\kappa = \frac{4k\_^2 \alpha}{\theta^2}$).

Chapter 6 - T Cell Activation I: Computational Investigations into the Origins of Short Term Biochemical Memory in T Cell Activation

Chapter 6.1. Introduction

The orchestration of the adaptive immune response is predicated on the integration of signals derived from peptide fragments that bear the molecular signature of an invading pathogen. T cells become activated by integrating signals derived from such peptides that are presented with proteins from the Major Histocompatibility complex (MHC) that are present on the surface of antigen presenting cells (APCs). Upon engagement with an APC, signal transduction is initiated by the interaction of the T cell Receptor (TCR) with MHC-peptide complexes. Successful signal integration results in numerous phenotypic outcomes and allows for the T cell to coordinate an appropriate immune response. In contrast, the failure of such priming processes leads to deleterious consequences such as autoimmunity.

The advent of two-photon imaging technologies has allowed for the study of real-time, in vivo T cell activation in lymph nodes in the presence of an antigenic challenge[140]. One consequence of such technological advancements is that recent imaging experiments have challenged the idea that the activation of naïve T cells requires prolonged continuous exposure from a single APC. These experiments suggest that, under certain conditions, T cells may integrate signals from short interrupted exposures to antigen presentation. For instance, in vivo mouse studies show that, during the activation process, there exists a stage where T-cells are involved in multiple transient interactions with many APCs[141].

In a parallel investigation into the nature of T cell signaling, a series of provocative *in vitro* experiments were conducted to specifically address the question of whether T-cells can integrate multiple interrupted signals and utilize the accumulation of

these signals for a biological response[142]. In their model system, signaling between Th1 T-cells and B-cell APCs was initiated in a collagen matrix. IFN-γ production along with other T-cell signaling markers such as calcium mobilization, ERK activity, and immunological synapse formation were monitored. Conjugation with APCs resulted in a sharp rise in calcium mobilization and ERK activity. In their system, IFN-γ production commences after roughly 30 minutes of active signaling through the cell-cell contact. The authors assessed whether T cells can integrate interrupted signals by introducing a reversible src family kinase inhibitor, PP2, after signaling had begun upon T cell – APC conjugation. This inhibitor is known to have a high selectivity towards Lck, a crucial src-family kinase responsible for triggering downstream pathways in T cells. Introduction of PP2 quickly abrogated both calcium mobilization and ERK activity, and the immunological synapse also rapidly disassembled. After a prescribed time interval of approximately 20 minutes, PP2 was washed out of the culture and signaling, as determined by calcium flux and ERK activation, resumed almost immediately; the immunological synapse also reassembles.

In addition, the authors report several other unexpected findings. They first demonstrate that thirty minutes of stimulation is initially insufficient for cytokine (IFN-γ, in this case) production. They then show that subsequent rounds of TCR signaling, after interruption of the signal, produce significant amounts of cytokine within thirty minutes. These results suggest that T-cells have the ability to integrate interrupted signals from multiple encounters with antigen and also suggests that T-cells can exhibit a "short-term" (the phrase "short-term" is used to distinguish from the 'long-term' development of the T cell memory phenotype) memory of past exposures to antigen in that the first exposure seems to prepare the T cell for subsequent

exposures to antigen. Fig. 6-1 gives a schematic of the main findings. Such a memory in this case persists for at least 20 minutes and moreover, there is also evidence to suggest , in other situations, that such a memory can persist for much longer periods of time[143]. The mechanistic origin of such a short-term memory is not understood although it is likely the result of the sustained activation of some signaling intermediate. Such a sustained activity can provide a means by which to integrate signals from transient cell-cell contacts and other types of interrupted signaling. Therefore, such an understanding may have direct applications related to the nature of transient versus stable T cell-APC contacts during T cell priming and activation.

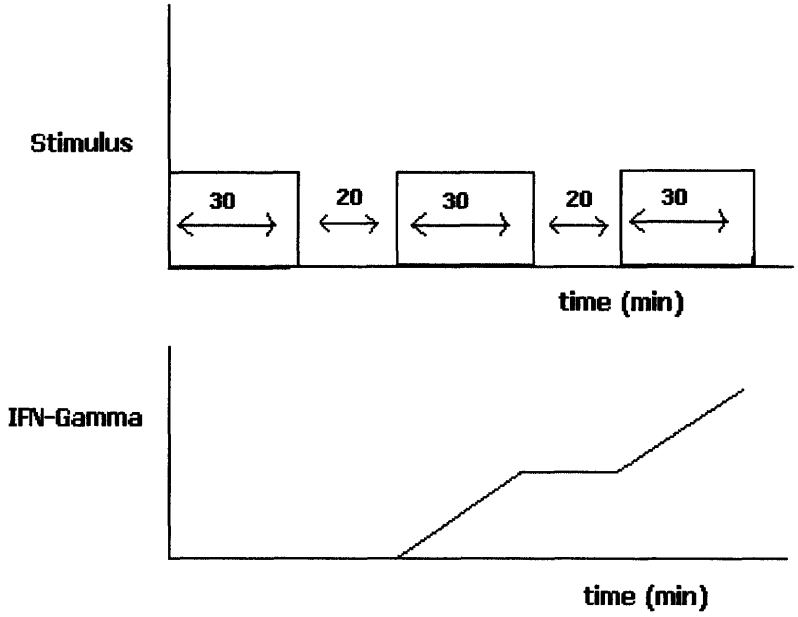


Figure 6-1. **Schematic of experimental results obtained demonstrating memory in signal integration**

a.) TCR mediated signal transduction proceeds for thirty minutes via a contact with APCs in a collagen matrix. After thirty minutes, signaling is aborted with the T-cell-APC contact intact. Following a period of approximately 20 mins, signaling is allowed to resume. The procedure is repeated periodically. b.) Cytokine (i.e. IFN-γ) production is measured. The first 30 minutes of signaling results in no IFN-γ production. Upon the next 30 minute round of TCR signaling, IFN-γ is produced. The final round of stimulation

> results in IFN-γ production as well. Curiously, the first round of signaling, while insufficient for IFN-γ production, prepares (i.e. establishes a memory) for more rapid IFN-γ production in subsequent rounds of stimulation.

The acquisition of 'short-term' memory is an emergent property that requires the coordination of cooperative, dynamic events involving many interacting cellular components. Thus, it is often difficult to intuitively assess how different molecules act in concert to exhibit collective, system-wide cellular behaviors such as signaling memory. In this regard, theoretical and computational studies[144] that involve mathematical modeling and allow for the systematic parsing of many dynamical processes have proven useful as a supplement to genetic, biochemical, and imaging experiments.

In the scope of this work, we present results from computer simulations of mathematical models that explore the physiological consequences of different mechanisms that can account for the biochemical memory observed in several experiments. We first briefly discuss some key observations from existing data in the literature and in particular, the experiment[142] that allowed us to construct our mathematical models. Next we present results from our modeling efforts that involve stochastic computer simulations of simplified versions of the signaling pathways leading to cytokine production.

Our goal was not to simulate every detail of the T-cell signaling network but rather to construct the simplest possible models that can investigate how the basic structure of several different molecular mechanisms could lead to the collective property of biochemical memory in T cell signaling. Other computational models have considered many elements of T cell activation such as how sensitivity to antigen is established[145,

146]. For example, 'digital' Erk responses that arise from the coupling of positive and negative feedback loops has been extensively studied as a model for signal integration at short time scales[145, 147]

Instead, our modeling efforts specifically aim to demonstrate how the first round of signaling can seemingly prepare the T cell for signal integration so that later exposures to antigen do not require a long lag time for cytokine production. Towards this end, we focus on events at later times that occur downstream of Erk activation. We derive from these models, testable predictions that can ascertain many aspects of such memory phenomena in T cell signal integration. The predictions that emerge from our calculations suggest several experiments that could further elucidate the possible mechanism for this 'short-term' biochemical memory. Finally, it is our hope that such calculations serve as a template for further quantification and modeling of memory phenomena and signal integration that are observed in T cell signaling.

Chapter 6.1.2 Possible sources for 'short-term' signaling memory

While initial signaling events, such as the mobilization of intracellular calcium and activation MAPK pathways, occur within minutes of the initial T cell/APC contact, at least thirty minutes of signaling is required for cytokine production. One hypothesis for the existence of this waiting period is that there is a time required for the accumulation of immediate early gene products (IEGs) such as Jun and Fos proteins which comprise the AP1 transcription factor complex. IEGs are synthesized de novo upon TCR signaling and their presence is a necessary condition for cytokine

production[148]. T cells then must undergo signaling for long time periods on the order of hours in order to become fully activated[149].

Studies on the duration of Erk signaling leading to the accumulation of IEGs suggest that a hyper-phosphorylated state of the IEG product Fos can remain active for long times[60, 150]. Furthermore, IEG products such as the family of Jun proteins have been observed to be active for long periods of time after the removal of TCR signals[151]. Hence, the activity of these transcription factors during periods of interrupted signaling could explain why the first round of signaling seems to prepare the T cell for cytokine production during later exposures to antigen. In this picture, the initial thirty minutes of signaling serve to accumulate IEG products that remain available for long times after the stimulus is removed. Then, for subsequent encounters with antigen, the activity of IEG products allows for faster cytokine production since this rate limiting step of the pathway is then bypassed. However, since all cytokine production ceases once TCR signaling is aborted, active IEGs alone can not be a sufficient condition for cytokine production. For instance, transcription factors, such as NFAT and NF-κb, derived from the activation of the calcium pathway and the PKC-theta pathway are other necessary conditions for cytokine production[152, 153].

One general way to generate sustained activity of signaling intermediates is to exploit positive feedback in signaling networks[154]. In a positive feedback loop, the end-product of the signaling pathway might up-regulate an activator of some upstream component of the pathway. Thus, once this activator is turned on, removal of the stimulus would not necessarily result in termination of the signal. Due to the autocatalytic nature of the feedback loop, provided that there is enough initial stimulation, active

signaling intermediates can be self-sustained even in the absence of stimuli. This phenomenon has been termed bistability, hysteresis, or effective irreversibility. There are other ways in which bistability could be generated; for example, disrupting the activity of an inhibitor of the pathway can also result in bistability.

Chapter 6.2. Results

Chapter 6.2.1 Mathematical models for biochemical memory

The first model that we studied is derived primarily from experimental results relating signal duration with the stability of IEG products[60]. Initial signaling events such as the activation of the calcium pathway and MAPK cascades occur within minutes of TCR-MHC engagement[155]. However, cytokine production does not commence immediately. As previously stated, one explanation for this waiting period is the time required for the accumulation of IEGs such as Jun and Fos proteins which comprise the AP-1 transcription factor complex. IEGs are synthesized *de novo* upon TCR signaling and are necessary for cytokine production[148]. Studies have shown that the duration over which activated Erk is maintained can be sensed by cFos, a protein product of IEGs; Erk can phosphorylate the Ser 362 and Ser 374 sites in cFos[60, 66]. This form of phosphorylated cFos is unstable; but, it is primed for additional phosphorylation by Erk. A DEF domain in cFos docks Erk, and primed cFos can be then phosphorylated at the Thr 325 and Thr 331 sites. This hyper-phosphorylated form of cFos apparently remains active for long times. One simple hypothesis could be that hyper-phosphorylated cFos is not subject to inactivation. The sustained activity of the hyper-phosphorylated form of cFos, and hence the transcription factor AP1, during periods of interrupted signaling then explains why the first round of signaling seems to prepare the T cell for cytokine

production during later exposures to antigen. In this model, the initial thirty minutes of signaling serve to accumulate IEG products which remain stable for long times following the removal of the stimulus. The accumulation could occur in a graded or switch-like manner as some have argued[156]. Then, for subsequent encounters with antigen, the activity of the IEG products allows for faster cytokine production since this rate limiting step of the pathway is removed from the signaling network.

However, since all cytokine production appears to cease once TCR signaling is aborted, active IEGs alone cannot be a sufficient condition for IFN gamma production. For example, calcium mobilization another necessary condition for cytokine production[153]. The calcium pathway ultimately leads to the activation of the transcription factor, NFAT, which then translocates into the nucleus. It has been documented that upon disruption of signaling, NFAT activity will quickly decay due to the presence of GSK – GSK phosphorylates NFAT which then signals NFAT to egress from nucleus[153]. This may be the reason why no cytokine production is observed once the signal is disrupted. If the initial period of signaling is short, then the stable form of cFos will not accumulate, and so T cells will be unable to add up signals from subsequent exposures to antigen. On the face of it, the observations discussed above support the speculation that the crux of the mechanism underlying the ability of T cells to integrate multiple interrupted signals is the creation of a hyper-phosphorylated stable form of cFos that is mediated by sustained Erk activation.

As it stands, this mechanism for signaling memory is not without difficulties. Degradation mechanisms that are mediated by various ubiquitin pathways have been shown to occur with members of the AP-1 complex such as Jun[157]. Moreover, turnover

of IEG products mediated by ubiquitin pathways can be very fast in cells and is known in many cases to occur at rates faster than the duration of interrupted signaling that was measured. Furthermore, it is not clear why a hyperphosphorylated form of cFos cannot be dephosphorylated by phosphatases on a time scale much faster than the time during which the signal has been disrupted. Recognition of this potential difficulty leads to one hypothesis: IEG products such as Fos and Jun are embedded within positive feedback loops that allow their activity to persist long after the stimulus has been removed. Due to the autocatalytic nature of the feedback loop, an active signaling intermediate may be self sustained, even in the presence of protein degradation, by the catalytic cycle that is initiated within the signaling cascade. This hypothesis led us to investigate the biological consequences of models involving both the presence and absence of feedback loops.

Chapter 6.2.2 Computer simulations of the signaling models

The three scenarios examined in detail, are depicted in Fig. 6-2a. In each scenario active IEG product (e.g. cFOS) serves as the biochemical memory. Because the detailed biochemical mechanism by which cFOS is activated is not entirely known, we considered two cases. In the first case, Fig. 6-2b, the kinetics of cFOS phosphorylation are determined by laws of mass action involving a simple linear reaction mechanism. In the second case, Fig. 2c, the stabilization of cFOS by ERK is achieved cooperatively—the degree of cooperativity is determined by a Hill function. Lastly, in Fig. 6-2d, we consider the case where the hyperphosphorylated state of cFOS is maintained by positive feedback. A description of the network topologies used in the simulations as well as the kinetic parameters is given in the methods section and in Table 1. The sensitivity of the

model to perturbations in the parameters used in the simulations is also discussed in the methods section.

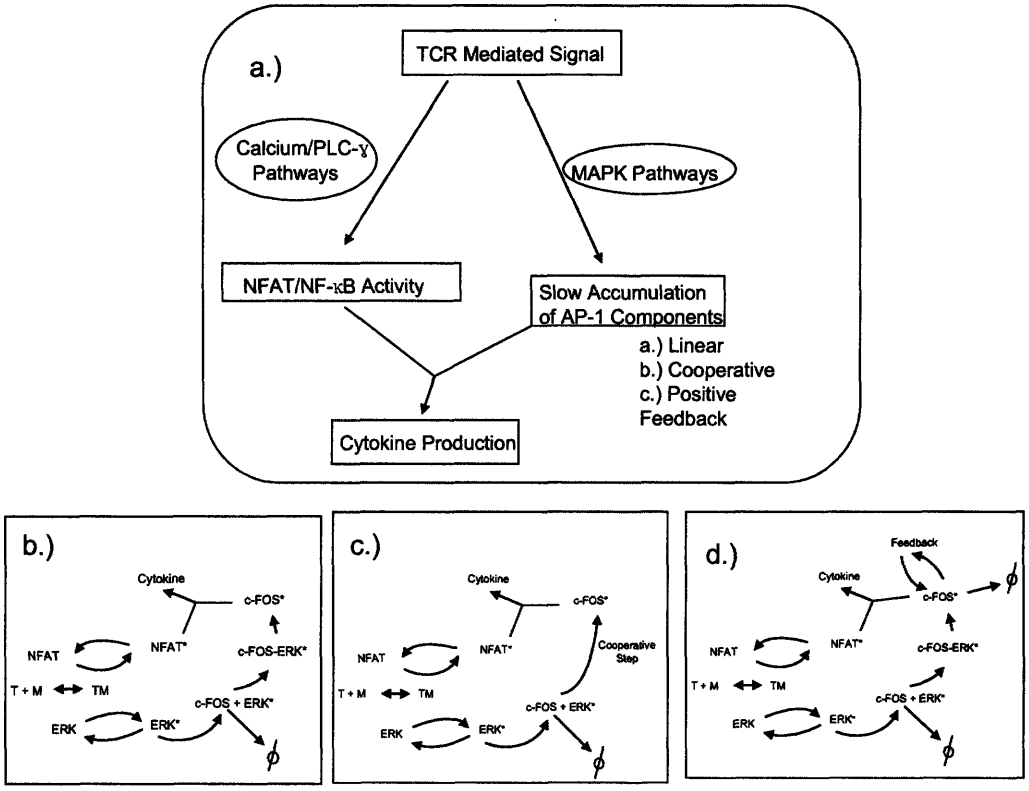


Figure 6-2. **Diagrams of the simplified signaling networks used in the computer simulations**

a.) An overall scheme for the signaling model to be simulated. Parallel pathways, whose activation occurs at different time scales, converge to produce cytokine. b-d.) Reaction schemes for each model, b.) linear c.) cooperative and d.) feedback induced models for persistent activity cFOS.

The calculations aim to mimic the experiments by periodically interrupting signaling by "inhibiting Lck" in the simulation for a period and then removing the "inhibitor". This is accomplished by disallowing any contribution of triggered T cell receptors to the activation of downstream pathways for a specified time interval. The

"strength" of the signal is determined by the duration of initial signaling, the number of agonist pMHC molecules, or the affinity of agonist molecules. Two general cases (defined in the methods) are studied: one in which the initial signal strength is large, and the other in which it is small; these values are defined more precisely in the context of each simulation.

Representative time courses are presented in Figs. 6-3 and 6-4. Consider first the behavior of calcium mobilization and its related transcriptional products (Figs. 6-3a, b). In the cases of low and high signal strengths, the activity of this pathway cycles approximately in phase with the cycling of the stimulus. This is because calcium mobilization and Erk activation are relatively fast in our model. For cases of weak stimulation, the signal cycles in phase with the duration of stimulation but is subject to large fluctuations (Fig. 6-3b) that may be interpreted as a less reliable signal.

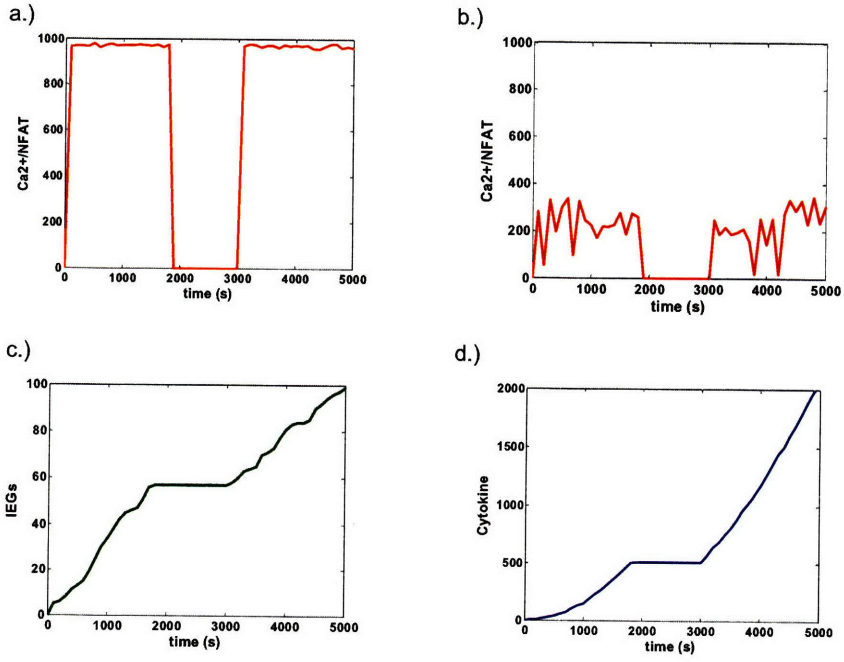


Figure 6-3. **Representative dynamics for cooperative and linear models**

a,b) Ca2+/NFAT dynamics. Under strong stimulation (a). Activity cycles roughly in phase with the duration of stimulation. Under weak stimulation (b), activity also cycles approximately in phase with the duration of signaling. However, such activity is less consistent than that observed in the case of strong stimulation and subject to large fluctuations. c,d.) Trajectories of active IEGs (e.g. cFOS) (c) and cytokine (d) for the case of cooperative cFOSp/Erkp dynamics in the presence of sufficiently strong stimulation. Other qualitatively similar cases are presented in the supporting online information.

In Figs. 6-3c,d, we focus our attention on the interaction of this pathway with the rest of the network—our results for the case where the stabilization of cFOS is cooperative are shown. In this case, the time courses for IEGs and cytokine production are very different from those showing Ca2+/NFAT activity. In Fig. 6-3c., IEGs slowly accumulate upon stimulation. Once the signal is disrupted, IEG accumulation halts and then resumes once the stimulus is reintroduced. Cytokine production (Fig. 6-3d.) then follows from the presence of IEGs; given a sufficient amount of IEG accumulation, cytokine is produced provided that the intermediates from the parallel pathway are active. On the contrary, for weak stimulation, there is, however, no IEG and cytokine production because the cooperative nature of the enzymatic reactions leads to the hyperphosphorylated stable form of cFos exhibiting an all or nothing response (data not shown). In the cooperative model, there is no response below a certain threshold of signal strength.

Now consider results from our computer simulations for the same model, but for the case where the stabilization of cFOS is not cooperative but rather occurs in a linear manner according to simple laws of mass action in the enzyme kinetics. Results from the

simulations show that there is no qualitative difference in the cases of strong and weak signal. Only the relative amounts of chemical species produced are different in the two cases. In this case, we observe a memory effect in the computer simulation irrespective of the strength of the signal(data not shown).

Finally, we observe the case where IEG products are embedded in an autocatalytic feedback loop (Fig. 6-4). For strong stimulation, we see production of stable IEG products that prepares for cytokine production at a later time(Fig. 6-4a). However, when the stimulus is disrupted, the amount of IEG decays to a steady value during the period of interruption. When stimulation is reinitiated, the amount of cFOS continues to grow monotonically and its activity contributes to the immediate production of cytokine(Fig. 6-4b).

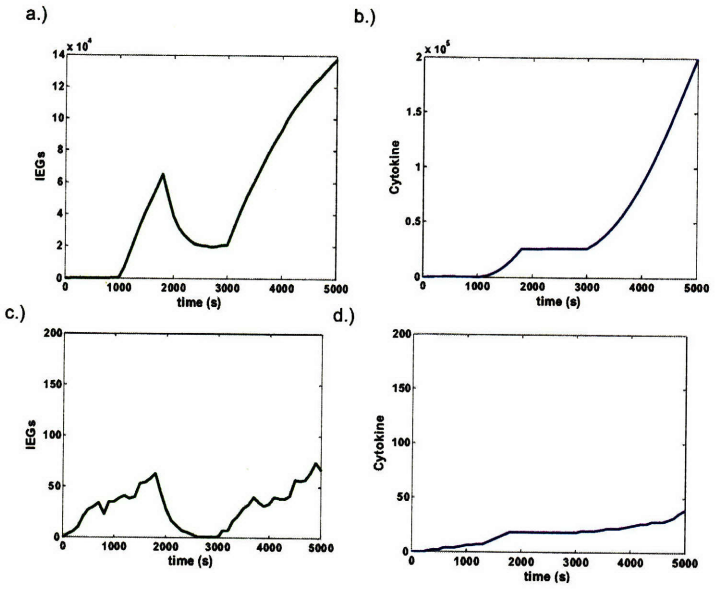


Figure 6-4. **Representative dynamics for models with positive feedback**

Scenarios involving strong (a,b) and weak (c,d) stimulation are considered. Representative timecourse for the activity of IEGs (a,c) and cytokine production (b,d) are shown. Memory effects are acquired in the case of strong stimulation. While weak stimulation

triggers IEG activity, such activity is effectively reset upon interruption of the signal.

Qualitative differences among the three models are further illustrated by monitoring the time evolution of probability distributions of pertinent signaling species. Such distributions are the analog to monitoring the statistics of the cell population. In Fig. 6-5, distributions of IEGs(Figs. 6-5a,b) and cytokines(Figs. 6-5c,d) produced at several time points are computed. Three time points are considered: at 30 minutes after the first round of signaling, at 50 minutes after the first period of interruption, and at 80 minutes after the second round of signaling.

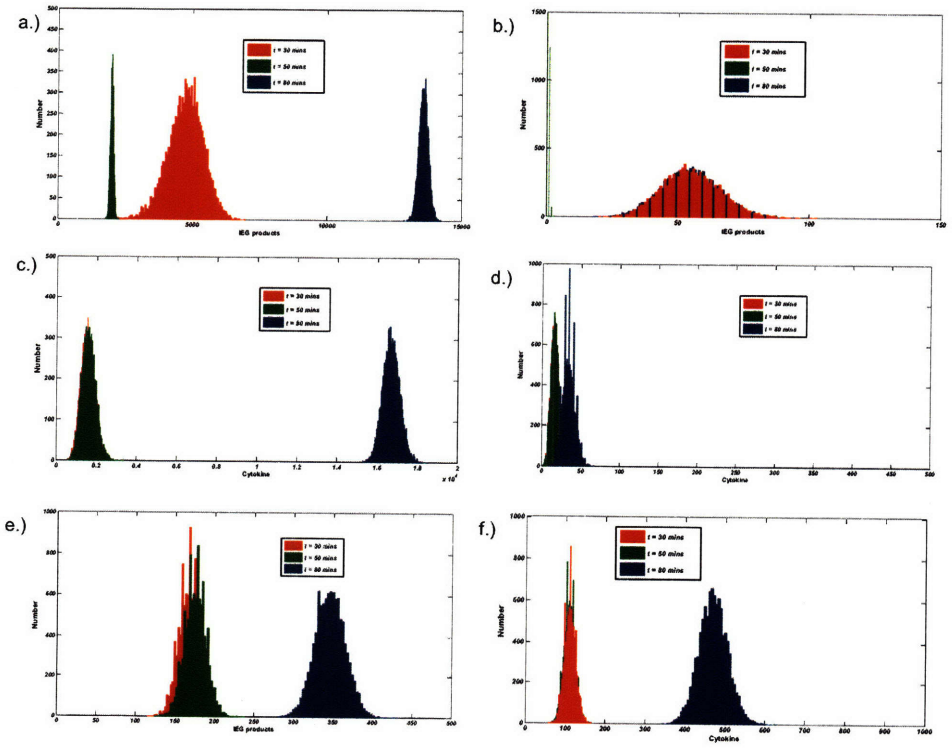


Figure 6-5. **Comparison of the distributions of active IEGs and cytokine production for different models**

a-d, results from feedback model. e,f, results from a linear model. Probability distributions are computed at three time points, t = 30 minutes (after first round of stimulation) red , t = 50 minutes (after first period of interrupted signaling) green , and t = 80 minutes (after the completion of the second round of signaling) blue. IEG products (a,b,e) and Cytokine production (c,d,f) are considered. In the presence of a feedback loop, two separate cases (strong (a,c) and weak (b,d) signal strength are analyzed.

In the presence of a feedback loop and sufficiently strong stimulation (Figs. 6-5a,c), we observe, at thirty minutes, a broadly peaked distribution centered on a large amount of IEGs (Fig. 6-5a). Little to no cytokine is produced at that time (Fig. 6-5c.). After signaling has been disrupted for 20 minutes, the simulated cell population of active IEGs shifts to the left and becomes sharply peaked. Now, at the end of the second round of signaling, the population remains sharply peaked and shifts markedly to the right and the number of IEGs and cytokines become greatly amplified(Figs. 6-5a,c). The feedback loop, in effect, allows for large signal amplification and reduces the amount of noise propagated in the signaling cascade(Figs. 6-5a,c).

For the case of weak stimulation(Figs. 6-5b,d), signal integration in the presence of a feedback loop shows very different qualitative behavior. After the first round of signaling, a broad distribution centered on a small amount of IEGs is observed (Fig. 6-5b). After the following twenty minutes of interrupted signaling, the entire population of IEGs decays to zero. The next round of signaling leaves the cell population identical to that which was observed at the end of the first round of signaling. Hence, the presence of a feedback loop along with rapid turnover of active signaling molecules suppresses all memory effects in the case of a sufficiently weak signal. These effects are also exhibited

in the distributions of cytokine production(Figs. 6-5c,d). In the first case where cFOS exhibits memory of the previous exposure to stimulation (Fig. 6-5c), we observe that the amount of cytokine produced is highly amplified in the second round of signaling whereas after the first round, cytokine production is barely detectable. In the case where no memory effect is observed, only minute amounts of cytokine production are observed – the amount of active IEGs is insufficient to produce significant quantities of cytokine.

In the absence of a feedback loop (Figs. 6-5e,f), the distributions of active cFOS (Fig. 6-5e) at 50 minutes are maintained at the same levels as were obtained after the first 30 minutes of signaling. This is because IEGs simply remain stable in these cases. After the second round of signaling, the amount of active cFOS about doubles and cFOS activity is not highly amplified after the second round of signaling. The distribution of cytokine production (Fig. 6-5f) follows from this result. A small amount of cytokine is produced in the first round of signaling and is proceeded in the second round by a much larger amount – this follows from the activity of cFOS that was acquired in the first of signaling. Similar qualitative behavior is seen in all other cases (cooperative and linear models, data not shown) except when enzymatic reactions are cooperative and an insufficient amount of stimulation leads to no activity. In such a case, no active cFOS, cytokine, and hence no memory is obtained; such effects arise from the switch-like or "all or none" nature of the signaling circuit for the cooperative model.

These results appear to suggest that experiments that probe the dose-dependent response of the signaling system would give qualitatively different results for each of the three models. The presence of a feedback loop would allow for an "all or nothing" memory effect and cytokine response, as would the presence of a cooperative "switch-

like" enzymatic stabilization mechanism. Alternatively, in the case of a model with linear enzymatic kinetics, a graded response would be predicted. A graded dose response is expected because there is no mechanism to give rise to a threshold in the signaling network. Such behaviors are displayed in Fig. 6-6 that gives a comparison of the dose response curves. Note that in the case of the feedback loop (Fig. 6-6c), a hysteresis is observed – the behavior of the dose response curve going forward is different from that obtained going backwards ("going backwards" refers to starting initially in the state with a large amount of active IEG). This sensitivity of signal output to the initial physiological conditions provides the source of memory.

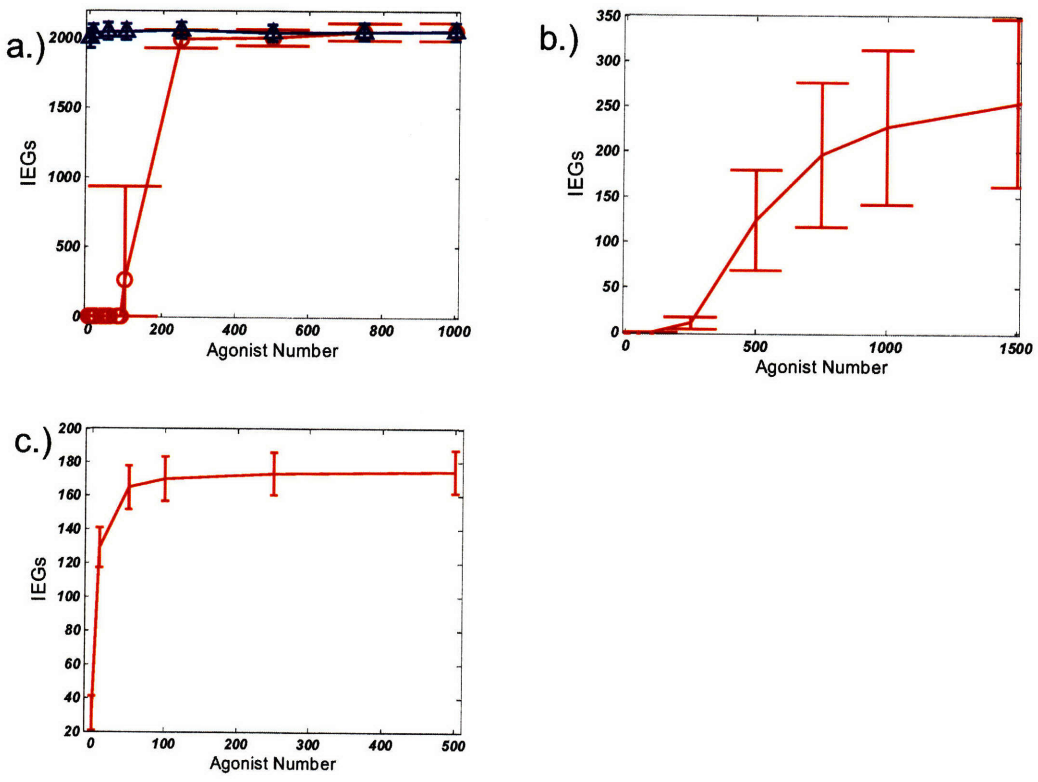


Figure 6-6. **Comparison of dose response curves**

Average values of active IEG products after the stimulus has been removed for 20 minutes (t = 50 minutes) a.) feedback regulation model. forward (red circles) and backward (blue triangles) curves are shown. Strong hysteresis is observed b.) cooperative reaction

model c.) linear, mass-action kinetics reaction model "error bars" are computed by considering the standard deviation—one measure of the magnitude of noise in the signaling process.

What might be the biological consequence of hysteretic effects present in the production of IEG products? We first consider the dependence of hysteresis on the strength of such a feedback loop. Such an effect in Fig. 6-6c implies that the persistence of memory effects in IEG accumulation can be made permanent. The backwards dose response curve in Fig. 6-6c indicates that, after twenty minutes of disrupted stimulus, such a memory effect for cytokine production will be apparent under all physiological conditions that may be realized during a subsequent round of signaling.

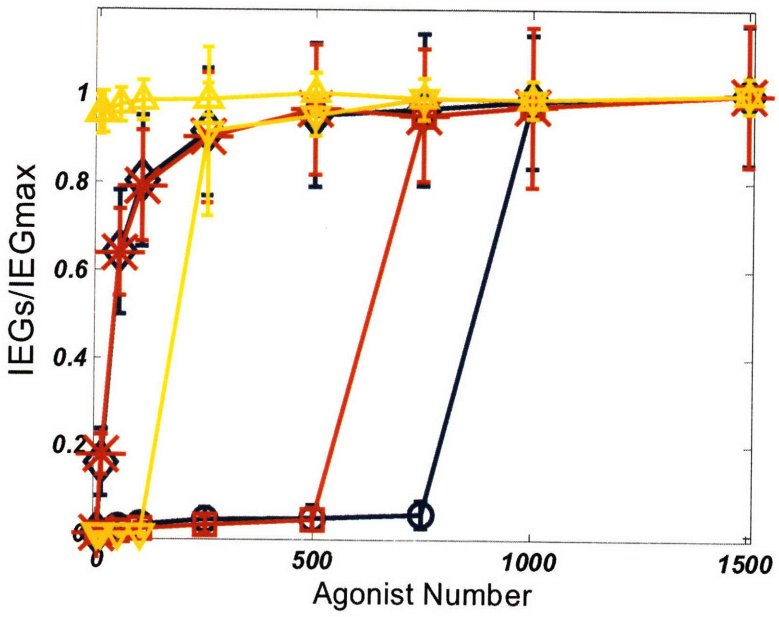


Figure 6-7. **Evaluation of the effects feedback strength**

Forward and backward dose response curves for varying feedback strengths, $\alpha = 1$ (blue), $\alpha = 2$ (red), and $\alpha = 5$ (yellow). Different markers correspond to the forward and backward dose response. At high feedback strengths, the response is irreversible. At low feedback strengths, the active state can reverted be back to the inactive state. Again, values are calculated at $t = 50$ minutes.

In contrast, Fig. 6-7. considers the effects of decreasing the feedback strength on the hysteresis in the signaling circuit. As the strength of the feedback loop, i.e. the value of α, decreases, the threshold signal strength required for acquisition of the memory effect increases and the curve markedly shifts to the right. Such a dependence of system behavior on the strength of the feedback could allow for some degree of plasticity in the response. For weaker feedback strengths, the dose response, while still retaining the switch-like characteristic, becomes reversible. Starting from the memory-competent state and decreasing signal strength, a point is reached at which the amount of active cFOS decays to zero (for $\alpha = 1, 2$ in Fig. 6-7). This implies that even if the first round of signaling is sufficient to induce such a memory with IEG products, a threshold amount of signal is required to achieve the memory effect. Therefore, cytokine production will only begin more quickly in subsequent rounds of signaling if the stimulation in that round is strong enough. It is interesting to speculate that such a control mechanism may serve to establish better specificity in the subsequent rounds of signaling.

Chapter 6.3. Discussion

Our computational analysis suggests specific experiments that could provide insights into the mechanisms that underlie the ability of T cells to integrate signals and retain a "memory" in the signaling process. The most significant experiments will be ones that monitor the stability of transcription factors in and out of the nucleus and determine whether individual activated molecules are stable or rather, constantly turning over when signal memory is exhibited. Signaling "memory" then can be assessed by the persistence of nuclear transcription factors after inhibition of the signaling pathway. Experiments with the Lck inhibitor PP2, in conjunction with immunofluorescence assays that make use of fluorescent secondary linked antibodies, can

monitor the nuclear translocation of the relevant transcription factors such as Fos, Jun, NF-κB, and NFAT upon disruption of TCR mediated signaling. These experiments will be essential to understanding the relevant transcription factors that enable short-term biochemical signaling memory.

Feedback loops are ubiquitous in T cell signaling[158] and evidence for bistability in signaling pathways has been shown in numerous cases[154]. Some of the most comprehensive studies involve studies of JNK signaling in Xenopus Oocytes[159]. These works demonstrate that the JNK pathway can both respond to stimuli in an all or none manner and exhibit all the features of a cascade involving strong positive feedback, including hysteresis. It is interesting to speculate that JNK signaling may exhibit similar features in T cells. The JNK cascade is involved in cytokine production, and exhibits many features of bistability. In CD8 T cells, the transcription factor c-JUN, a product of the JNK cascade, remains active for up to 24 hours after the stimulus has been removed[151]. However, JNK activity has recently been shown, in one case, to be unnecessary for IFN-γ production. It is also possible that signaling leading to the production of protein products of IEGs (e.g., cFos) is embedded in a positive feedback loop.

Furthermore, one prediction obtained from this model suggests that one could distinguish between the two possible mechanisms for sustained activity by carrying out photobleaching experiments using GFP constructs of the relevant transcription factors. If signaling memory is due to a long half-life of the signaling intermediate, photobleaching will largely eradicate the ability to observe nuclear localization of the transcription. If, on the other hand, signaling memory is due to existence of positive feedback in the signaling circuit, then one would expect that nuclear fluorescence will rapidly recover after photobleaching.

These experiments will first help to first determine whether AP1 is the transcription factor that is the source of biochemical memory, and should allow one to discriminate between the models of biochemical memory we have studied *in silico*. Should it be found that a particular model where individual activated molecules are rendered stable for long times is appropriate, the importance of cooperative enzymatic modifications can be determined by measurements of dose-response curves for the amount of activated cFos as a function of signal strength, which could be modulated by the amount or quality of the agonist pMHC as well as the duration of the initial signal.

Directly observable predictions about how the kinetics and strength of TCR signaling affects signaling memory could be tested. Since a model involving either a cooperative mechanism or a feedback loop predicts a threshold for a memory effect in signal transduction, the strength of signal will determine whether or not T cells can integrate signals during multiple exposures to antigen. These models propose that there exists some crossover between weak and strong agonists, short and long durations of TCR signaling, and concentrations of low and high numbers of agonist; this crossover will determine whether or not a lag time is required for cytokine production during subsequent rounds TCR signaling after the signal has been disrupted.

Signaling memory implies the persistence of sustained activity of some biophysical or molecular signaling intermediate even after signaling is interrupted by removing the stimulus. Such an intermediate could be the persistence of spatially clustered signaling components (e.g. as the formation of microclusters)[160, 161] or the presence of compartmentalized signaling components[65] whose prior assembly constituted a rate limiting step. However, many of these processes are actin-mediated[160] and likely rapidly aborted in the absence of a signal. In light of these difficulties, it seems that a model invoking positive feedback is a plausible explanation

for the molecular origins of memory in T-cell signal integration. Such a model is desirable on several bases; it provides noise reduction, plasticity in threshold tuning, precise control of signal amplitude and timing, and potentially useful hysteretic effects in the acquisition of such a signaling memory. The other models lack most, if not all, of these features. However, such memory effects in the form of spatial localization or perhaps time delays can not be excluded at this time.

In summary, we have explored, *in silico*, several molecular models which can explain the mechanism of biochemical memory in T cell signaling and activation. Each model involves the sustained activation of a certain transcription factor in the presence of disrupted signaling. Furthermore, our computer simulations make several predictions that we have briefly outlined.

It is our hope that this work will serve to motivate as well as guide future experimentation into mechanisms underlying biochemical memory in T-cell signaling and activation. Once these mechanisms are better understood, further elaboration on the details of our computer models will be necessary to provide a better quantitative analysis of the mechanism governing the memory phenomenon. Also, it will be important to address in the near future how signaling memory at the cellular level functions in the context of T-cell activation in vivo where T-cell migratory patterns in lymph nodes are important in controlling the overall outcome of the physiological response. Integration of a more detailed computational model of the signaling pathways that maintain short-term memory with a computational model for T-cell trafficking in lymph nodes will be essential for understanding this problem. A model of this sort can then be used in conjunction with two-photon imaging experiments *in vivo* along with genetic and biochemical experiments to investigate the underlying mechanisms in T-cell activation across multiple length

and time scales, from the molecular features governing the dynamics of signaling pathways to the clearance of infection occurring at higher levels of biological organization.

Chapter 6.4. Methods

Chapter 6.4.1 Signaling reactions

The signaling models that we chose to simulate consist of the following half reactions and the basic set of molecular processes common to each of the three models is as follows (cFOS is taken to be the example of the Immediate Early Gene product):

$$T + M \rightarrow TM \quad (6\text{-}1)$$
$$TM \rightarrow T + M \quad (6\text{-}2)$$
$$NFAT + TM \rightarrow NFAT^* + TM \quad (6\text{-}3)$$
$$NFAT^* \rightarrow NFAT \quad (6\text{-}4)$$
$$ERK + TM \rightarrow ERK^* + TM \quad (6\text{-}5)$$
$$ERK^* \rightarrow ERK \quad (6\text{-}6)$$
$$ERK^* + GENE \rightarrow cFOS + ERK^* \quad (6\text{-}7)$$
$$cFOS \rightarrow NULL \quad (6\text{-}8)$$
$$cFOS^* + NFAT^* \rightarrow Cytokine \quad (6\text{-}9)$$

Reactions determining cFOS* accumulation are shown for the linear (a), cooperative (b), and feedback (c) models.

a.) Linear model

$$cFOS + ERK^* \rightarrow cFOS\text{-}ERK^* \quad (6\text{-}10a)$$
$$cFOS\text{-}ERK^* \rightarrow cFOS + ERK^* \quad (6\text{-}11a)$$
$$cFOS\text{-}ERK^* \rightarrow cFOS^* + ERK^* \quad (6\text{-}12a)$$

b.) Cooperative model

$$cFOS + ERK^* \rightarrow cFOS^* + ERK^* \quad (6\text{-}10b)$$

c.) Feedback model

$$cFOS + ERK^* \rightarrow cFOS^* + ERK^* \quad (6\text{-}10c)$$
$$cFOS^* \rightarrow cFOS \quad (6\text{-}11c)$$
$$cFOS^* + GENE \rightarrow cFOS^* + FEEDBACK \quad (6\text{-}12c)$$
$$FEEDBACK \rightarrow NULL \quad (6\text{-}13c)$$
$$FEEDBACK + GENE \rightarrow cFOS^* + FEEDBACK \quad (6\text{-}14c)$$

The parameters used in the Monte Carlo simulations are in Table 1.

We simulated these models by solving a master equation[138],

$$\frac{\partial}{\partial t}P\left(\vec{n},t\right)=\sum_{\vec{n}'}W_{\vec{n}'\vec{n}}P\left(\vec{n}',t\right)-\sum_{\vec{n}'}W_{\vec{n}\vec{n}'}P\left(\vec{n},t\right)\ , \qquad (6\text{-}15)$$

whose solution gives the time evolution of the probability distribution for the system of chemical species to be in state $\vec{n}$. $\vec{n}$ is a vector whose components give the number of molecules of each molecular species. $W_{\vec{n}\vec{n}'}$ gives the transition probability per unit time for the transition of $\vec{n}$ to $\vec{n}'$. We used the standard stochastic simulation algorithm developed by Gillespie for solving master equations involving chemical reactions [162]. We constructed $W_{\vec{n}\vec{n}'}$ by first considering "mass action" kinetics that are determined by the topology of the reaction network corresponding to each signaling model. For the more complicated reaction mechanisms that we invoked to model cooperativity and feedback, we instead use the following unit-time transition probabilities,

$$W^{COOP}_{ncFOS^*\rightarrow(n+1)cFOS^*}=\frac{\alpha[cFOS][ERK^*]^H}{\beta^H+[ERK^*]^H} \qquad (6\text{-}16)$$

and

$$W^{FEED}_{ncFOS^*\rightarrow(n+1)cFOS^*}=\frac{\alpha[cFOS^*]^H}{\beta^H+[cFOS^*]^H} \qquad (6\text{-}17)$$

respectively. α and β are adjustable parameters that determine the strength of the nonlinear interaction. H determines the degree of cooperativity. Distributions were compiled from simulations of 10,000 statistically independent trajectories for each case presented. When plotting average behavior, error bars were obtained from simulations of 1000 trajectories. All code was written in ANSI C and compiled with the gnu C compiler, GCC.

The set of kinetic parameters used in the simulations is shown in Table 1. It is important to note that the simple signaling models we presented are not designed to quantitatively reproduce or fit experimental data; rather, their purpose is an attempt to lend deeper insight into the nature of such signaling mechanisms and generate useful predictions. However, our choice of parameters is not arbitrary; parameters were first estimated and constrained by way of a careful analysis of the important, experimentally measured time scales in the signaling process. Then, sensitivity of these parameters to the various mechanisms in question was studied.

Table 1. Parameters used in the Monte Carlo simulations

Reaction	Rate Constant	Linear Model		Feedback Model	
1	0.0000001	10a	0.1	10c	0.1
2	0.019	11a	0.1	11c	0.1
3	0.01	12a	0.1	12c	0.1
4	0.1			13c	0.1
5	0.01	Cooperative Model		α	100
6	0.1	α	0.01	β	600
7	0.0005	β	600	H	10
8	0.1	H	10		
9	0.000001				

Table 1. **Parameters used in the Monte Carlo simulations**

Chapter 6.4.2 Sensitivity of model to changes in biophysical parameters

The mathematical models of cell signaling that we analyzed are comprised of several modular components. Therefore, the sensitivity of the qualitative results of our models to the choices of kinetic parameters may best be understood by considering the

key competing time scales, $\{\tau_{sig}, \tau_{p1}, \tau_{p2}, \tau_{mem}, \tau_{cyt}\}$, that emerge in the modular network architecture that we constructed.

τ_{sig} is the time scale for signals derived from TCR-MHC to propagate to downstream messenger pathways. τ_{sig} emerges from kinetic constants and initial concentrations in reactions (6-1) and (6-2). τ_{sig} then, is a measure of the overall signal strength, which can be varied by adjusting the agonist concentration. For example, high strength (1000 pMHC molecules) and low signal strength (10 pMHC molecules) as well as long and short durations of signal map onto a value of τ_{sig}. τ_{p1} and τ_{p2} are the characteristic time scales involved in activating the two parallel messenger pathways in our model. τ_{p1} is the time scale to activate the fast pathway (e.g. Ca2+ Mobilization and active NFAT). τ_{p2} is the time scale required to activate the other pathway that leads to the synthesis of unstable IEG products. τ_{mem} is the time needed to establish a biochemical memory in the signaling circuit. A model assumption is that $(\tau_{p1} \sim \tau_{p2}) << \tau_{mem}$. If this were not the case (i.e. $(\tau_{p1} \sim \tau_{p2}) > \tau_{mem}$) then subsequent rounds of signaling would not quickly produce cytokine. Thus, τ_{p1} and τ_{p2} as well as the the time scale for cytokine production τ_{cyt} then limits the speed at which productive signaling can recover from interrupted stimulation. A mechanism involving the stabilization of IEGs as a source of memory requires that τ_{mem} be large – at least on the order of minutes.

Parameters from each model contributing to τ_{mem} (i.e. those in reactions 6-10a-c, 6-11a,c ,6-12a-14c) were varied and results are either presented in the main text or are discussed below. For the linear model, τ_{mem} changes in response to the kinetic

parameters in reactions 6-10a, 6-11a, 6-12a and subsequently determines the amount of stable IEG but does not affect qualitative findings. For the cooperative model, parameters in reaction 10b then control the positioning and sharpness of the threshold. α controls the amount of stable IEG that can be obtained. The behavior of the parameters involved in the feedback loop is discussed in the results section. Finally, τ_{cyt} is a time required for cytokine production once signaling intermediates (available IEGs and Transcription factors come that from the other pathway) are available. Changing τ_{cyt} then results in changing the amount of cytokine produced in the simulation in a monotonic fashion.

For the most part, we found that many of the qualitative results obtained from our models are robust to large (greater than 10-fold) variations in key individual parameters. However, two key parameters in our model could potentially change the qualitative results of our computer simulations; these parameters are the rate constant for the de novo synthesis of IEGs (reaction 6-7) and the rate constant for IEG decay (reaction 6-8). This is because, in our model, there exists a competition between the synthesis of cFOS and its decay. Upon varying these rates, we find that our key result can be sensitive to the rate of cFOS production. If cFOS synthesis is too slow, then no stabilized cFOS will be present once the stimulus is removed at t = 30 minutes in our simulations.

Chapter 7 - T Cell Activation II: Signal Integration at the Immunological Synapse and the Effects of Altered Peptide Ligands

Chapter 7.1. Introduction

The activation of T lymphocytes (T cells) is stimulated by the binding of the T cell receptor (TCR) to its ligand, short peptides (p) bound to major histocompatibility (MHC) gene products. Initially it was thought that each TCR was specific for a single pMHC molecule, but it is now clear that a variety of related peptides can stimulate signaling through a particular TCR. The specific sequence of a peptide determines its stimulatory ability and the resulting functional outcome.

Altered peptide ligands (APLs), generated by mutating residues of the wild-type agonist peptide, have been used to investigate the role of antigen quality on T cell activation as well as mechanisms underlying TCR signaling (Evavold et al., 1993). APLs can enhance or reduce the stimulatory ability compared to the wild type (WT) agonist peptide. The stimulatory potency of a peptide has been correlated with a variety of parameters that include the dissociation rate of the TCR/pMHC complex, the ability to downregulate TCRs, the ability to form an immunological synapse (IS), and the ability to generate fully phosphorylated TCR-ζ chains (Bachmann et al., 1997; Grakoui et al., 1999; Itoh et al., 1999; Kersh et al., 1998a; Madrenas et al., 1995; Sloan-Lancaster et al., 1994; Valitutti et al., 1995). The importance of the density of antigenic pMHC molecules expressed on an APC has also been noted (Gonzalez et al., 2005). How each of these parameters is linked to biophysical features of the TCR-pMHC interaction, stimulatory potency, and T cell signaling is not clearly understood. In this paper, we try to parse some of these relationships by studying the relationship between the IS, the structure formed at the T cell-APC contact site, with signaling using the AND TCR transgenic T

cells (which recognize moth cytochrome C (MCC) peptide presented by I-E^k (Kaye et al., 1989) with a series of APLs.

The IS refers to the organization of membrane proteins and cytosolic molecules that forms at the junction between a T cell and an antigen presenting cell (APC), (Grakoui et al., 1999; Monks et al., 1998). In some ISs, clustering of TCR and pMHC ligands in the center of the contact area occurs (called the cSMAC). In spite of numerous studies, the function of the cSMAC remains elusive (Campi et al., 2005; Lee et al., 2003; Lin et al., 2005; Yokosuka et al., 2005).

We combined *in vitro* experiments with computational studies to examine how TCR signaling may be influenced by the biophysical parameters characterizing TCR-pMHC binding and cSMAC formation. The computational results show that, for pMHC ligands that bind the TCR with a sufficiently long half-life, the primary function of the cSMAC is to enhance TCR downregulation. One surprising consequence of this effect is our finding that a weak pMHC ligand that does not induce cSMAC formation results in enhanced T cell proliferation compared to the wild-type agonist. To prove this, we used the ability of NKG2D, a receptor expressed on NK cells and some CD8 T cells, which can enhance cSMAC formation regardless of antigen quality [163], and found that cSMAC formation inhibited the stimulatory capacity of the weak peptide agonist. Our computational studies, however, also indicated that the function of the cSMAC for T cells may not always be inhibitory, and suggest future experimentation that could help resolve current debates.

Chapter 7.2. Results

Chapter 7.2.1 Integrated signal depends upon antigen quality in a complicated manner

Seemingly paradoxical results show, for example, that the K99A peptide does not induce efficient cSMAC formation but elicits a greater proliferative response compared to the WT peptide (which induces a cSMAC) suggested a complex interplay between competing phenomena that was difficult to intuit. We therefore undertook a computational study to gain further insights.

Since T cell activation requires prolonged signaling, we assumed that the integrated amount of signaling over time would correlate with proliferation. How the integrated amount of signal depends upon the on and off-rates characterizing TCR-pMHC binding and cSMAC formation was then analyzed. Because the role of the cSMAC in regulating TCR degradation is currently controversial, we examined the consequences of two different hypotheses. In the first hypothesis, no special assumption was made regarding where receptor degradation can occur. Receptor-ligand binding, downstream signaling reactions, and receptor degradation were allowed to occur wherever the appropriate molecules mediating specific chemical reactions encountered each other. In the second hypothesis, following suggestions in the literature (Campi et al., 2005; Yokosuka et al., 2005), it was a-priori assumed that receptor degradation could only occur in the cSMAC.

Computational investigation of either hypothesis requires a model for TCR signaling. The TCR signaling model that we studied (described in the Methods) is a

modification of our previous model (Lee et al., 2003). In the current model, we incorporated a more realistic mechanism for TCR degradation incorporating ubiquitination by allowing for the recruitment of the E3 ubiquitin ligase, Cbl, to phosphorylated ZAP-70 and CD3ζ molecules. The interaction resulted in ubiquitination and only ubiquitinated receptors were subject to degradation. Thus, degradation is naturally linked to the level and concentration of phosphorylated substrates. In the first model, Cbl was allowed to interact with receptors at any area in the simulation. In the second hypothesis, it was assumed that these processes can only occur in the cSMAC.

Chapter 7.2.2 Relationships between half-life of TCR-pMHC complexes, cSMAC formation, and integrated signal.

Calculations were first carried out to examine the effects of k_{off} and cSMAC formation on the integrated signal, fixing the on-rate for TCR-pMHC binding. We studied two situations, one where a cSMAC was allowed to form regardless of the value of k_{off}, and the other where the cSMAC never forms. Consider first representative data for the first hypothesis where no special assumption was made about where receptor degradation can occur (Fig. 1A). These results showed that, for ligands that bind the TCR with a long half-life, cSMAC formation led to a reduced amount of integrated signal. On the other hand, for ligands that bind the TCR with a short half-life, cSMAC formation is predicted to lead to an increase in the total integrated signal.

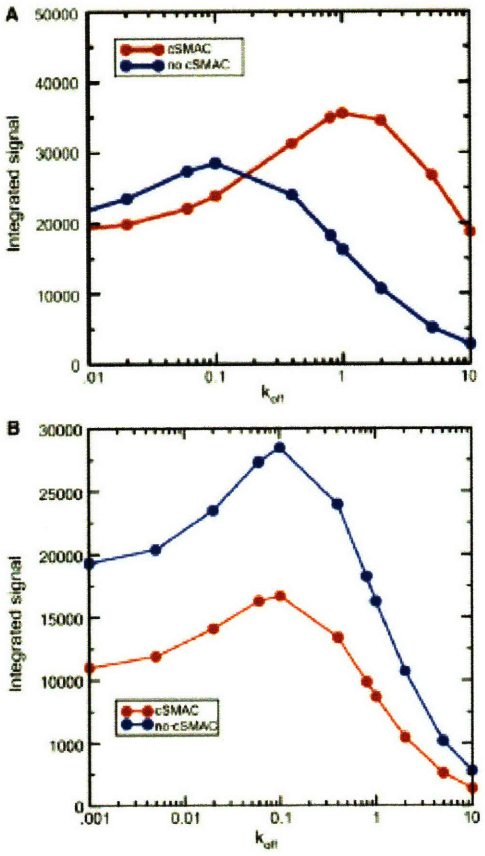


Figure 7-1. **Dependence of integrated signal on antigen quality**

(A) Data for calculations are compared for two cases, one where a cSMAC (red circles) is allowed to form regardless of the value of k_{off} and the other where no cSMAC (blue circles) is ever present. For small values of k_{off}, cSMAC formation inhibited the total amount of integrated signal, while the opposite was true for ligands that bind TCR weakly. These calculations were carried out with a pMHC density of 1 molecule/$(\mu m)^2$. Higher pMHC densities (e.g., 10 molecules/$(\mu m)^2$) did not change the qualitative behavior. These calculations were carried out for a value (Li et al., 2004) of k_{on} equal to 2200 M^{-1} s^{-1} (B) Integrated signal from a model (red line) where the cSMAC serves as a site for degradation only is compared with the situation where there is no cSMAC formation (blue line). The signal is always higher when there is no cSMAC formation and there is no "neutral" ligand as in Fig. 1A.

Analyses of our results showed that, for ligands with long half-lives (small k_{off}), receptors are triggered efficiently without need for clustering in the cSMAC. However, we found that, for these ligands, cSMAC formation increases the rate of degradation. Our analysis showed that this is because concentrating receptors in one spatial location (cSMAC) enhances the ability of Cbl to ubiquitinate substrates. Thus, the lower total integrated signal seen for strong ligands when cSMACs can form is explained by increased receptor downregulation.

At larger values of k_{off} (half-life is short), this was reversed.
Since weakly binding ligands did not trigger TCR efficiently, clustering of receptors, ligands, and kinases in the cSMAC aids in the generation of phosphorylated receptors and activated signaling intermediates because the enhanced concentration increased receptor occupancy and rates of phosphorylation. For these qualities of ligand, the integrated signal was distinctly smaller when the cSMAC did not form. Because these ligands induce lower levels of phosphorylation, the effects of cSMAC formation on signaling dominated over the effects of cSMAC formation on receptor downregulation. Importantly, the two curves in Fig. 1A (with and without cSMAC formation) intersect at a particular value of k_{off} where the advantages and disadvantages of cSMAC formation are balanced.

A critical issue in the interpretation of these simulations for the first hypothesis is the threshold value of k_{off} required to stimulate cSMAC formation (Grakoui et al., 1999; Sumen et al., 2004) and the relationship between this threshold and the theoretically point at which the two integrated signaling curves intersect. If the value of k_{off} that triggers cSMAC formation does not correspond to the intersection point, the amount of integrated

signal will exhibit a discontinuous change when T cells are stimulated by a ligand that is slightly weaker than that required for efficient cSMAC formation. For example, if the threshold value of k_{off} that triggers cSMAC formation is higher (to the right) than the intersection point, the amount of integrated signal will precipitously decrease for a ligand that has a value of k_{off} just above this threshold required for forming cSMACs. If the threshold value of k_{off} that triggers cSMAC formation is smaller (to the left) of the intersection point, the amount of integrated signal will abruptly increase for ligands whose binding to TCR has a value of k_{off} just above this threshold. Our experimental results showing that K99A elicits a stronger proliferative response than MCC, but does not form cSMACs efficiently, suggests that if the first model is valid, the threshold for cSMAC formation lies to the left of the intersection point for AND T cells. It also suggests that if T cells stimulated by K99A could be manipulated to efficiently form a cSMAC, the integrated signal would be lower.

Calculations for the second hypothesis, where it is assumed that receptor downregulation can only occur in the cSMAC are shown in Fig. 7-1B. In contrast to Fig. 7-1A, the curves with and without cSMAC formation do not cross at any value of k_{off}. In this model, cSMAC formation is predicted to reduce the amount of integrated signal for all antigen qualities. Thus, this hypothesis is also consistent with our experimental data on K99A and is independent of the threshold value of peptide quality required to form the cSMAC. Differences between the results in Figs. 7-1A and 7-1B suggest a way to distinguish between the two hypotheses. For example, if a sufficiently weak ligand that normally cannot stimulate cSMAC formation could be coerced to form a cSMAC and the stimulatory potency increased then the first hypothesis would be correct.

Chapter 7.3 Discussion

How different parameters determine the ability of a pMHC molecule to stimulate T cell activation remains unclear. Many studies (Kalergis et al., 2001; Kersh et al., 1998b) have suggested that the off-rate characterizing the binding of pMHC ligands to TCR correlates well with its ability to stimulate T cell activation. However, several anomalies have been reported that raise questions about whether this is the only measure of antigen quality. Thus, other parameters and phenomena such as the change in heat capacity upon receptor-ligand binding (Krosgaard et al., 2003; Qi et al., 2006), serial triggering, synergy between endogenous and agonist ligands (Wulfing et al., 2002; Krogsgaard et al., 2005; Li et al., 2004), and co-stimulatory interactions (Markiewicz et al., 2005; Purtic et al., 2005; Wulfing et al., 2002) have been implicated in modulating how T cells respond to antigen. Our studies combined *in vitro* and *in silico* investigations to examine the interplay between signaling, peptide quality and cSMAC formation.

Our results emphasize that the stimulatory potency of a pMHC ligand depends upon many factors, and that agonist quality does not correlate with any one variable. A dramatic example of this is provided by our experimental data for the stimulatory potency of the K99A peptide for AND T cells. This ligand binds the AND TCR with a larger value of the off-rate compared to the WT MCC and does not induce cSMAC formation. Yet, it elicited a greater proliferative response. This result raises the question as to whether other APLs classified only by proliferation assays as superagonists might actually be weaker binding ligands that do not induce cSMAC formation resulting in a higher total integrated signal.

Chapter 7.3.1 The role of the cSMAC

One idea emerging from our studies is that the dependence of cSMAC formation on TCR-pMHC half-life can modulate how the integrated amount of TCR signaling (and proliferative response) varies with receptor-ligand binding parameters. Our calculations showed that, for some ligands that bind TCR sufficiently strongly, weaker ligands that do not induce cSMAC formation should induce greater proliferative responses. This explanation for our experimental data on AND T cells stimulated by the K99A peptide are clearly supported by our results showing the reduced stimulatory potency of this peptide upon forcing cSMAC formation by expression of NKG2D. The results of this experiment are especially striking as NKG2D is thought to function as a co-stimulator to enhance signaling through the TCR.

We should also note that the results of the experiment using NKG2D are consistent with both hypotheses whose consequences we examined using in silico modeling. Both Figs. 6A and B show that, if the half-life for TCR-pMHC complexes is relatively long, a ligand that cannot induce cSMAC formation will result in a larger integrated signal. This result is also consistent with our previously reported experimental results that showed that disruption of cSMAC formation due to CD2AP deficiency resulted in prolonged signaling and a more robust response when T cells were stimulated by a strong agonist (Lee et al., 2003).

Chapter 7.3.2 The cSMAC and degradation

Recent work using single molecule imaging of T cells interacting with artificial lipid bilayers loaded with strong agonists suggest that signaling occurs mainly at the periphery of the contact area in microclusters (Campi et al., 2005; Yokosuka et al., 2005).

Consistent with this result, our computer simulations showed that, for high affinity agonists, TCRs are efficiently triggered in the periphery without need for clustering of molecules in the cSMAC. Our computational model, however, does not explicitly treat microclusters since it is a coarse-grained model that describes species in terms of concentrations. However, we expect the qualitative results reported by us to be similar if we were to carry out a simulation where individual receptor-ligand interactions and their clusters were treated explicitly. This is because, for strong agonists, we find that receptor triggering occurs efficiently in the pSMAC in our coarse-grained calculations just as is posited to occur in microclusters.

The paucity of phosphotyrosine staining in the cSMAC when T cells are stimulated by strong agonists has been interpreted to signify that the cSMAC functions only as a site of degradation (Mossman et al., 2005). Our computational results suggest that these past studies and the results reported by us in the present paper cannot unequivocally reach this conclusion because the situation could be different for weak ligands that are yet to be studied.

The major difference between the two hypotheses that we examined is manifested in the response to a peptide with a much shorter half-life than any we have tested. For peptides with very short half-lives, the first hypothesis (wherein no assumption is made about where degradation may occur) predicts that if cSMAC formation could occur, it would result in enhanced recognition of weak peptides. In contrast, the second hypothesis (where it is posited that degradation can only occur in the cSMAC) predicts that cSMAC formation will always weaken signaling. Our efforts to distinguish between

these two hypotheses in this way are currently constrained by the fact that no adequate weak peptides are known for the AND TCR system.

It could be argued that the weak peptides that would distinguish between these two models are not physiologically relevant as they would probably be too weak to activate T cells. However, it has recently become clear that endogenous peptides that are unable to stimulate T cell activation on their own are likely to play significant roles in TCR triggering [164]. The presence or absence of cSMACs could potentially determine the stimulatory capacity of such endogenous peptides at late time points in T cell signaling. This also raises important issues about *in vivo* situations where mixtures of both weak and strong ligands are present. Furthermore, the function of the cSMAC for less sensitive naïve T cells requires careful examination.

Our understanding of the function of the cSMAC is still evolving. We hope that continued synergy between modeling and experiments will help unravel this enigma.

Chapter 7.4. Methods

A continuum model that parallels and complements our previous stochastic models (Lee et al., 2003) was developed This enables us to readily examine large systems under diverse conditions and exhaustively examine the sensitivity of the model to values of unknown parameters. However, stochastic models (Lee et al., 2003; Li et al. 2004) are necessary if the copy number of certain species is very small. Since most *in vitro* experiments are carried out with a rather high antigen dose (e.g., concentration of agonist peptides loaded on supported lipid bilayers is 1 - 100 molecules/μm^2 (Grakoui et al., 1999))., use of the continuum model is appropriate. Varying the densities of various species led to qualitatively similar results over a range of values, and the effects of

fluctuations for low pMHC densities and small values of the off-rate, were not large for the conditions of interest in this paper.

In the model, the membrane associated and cytosolic molecules undergo biochemical reactions according to the signaling cascade shown in Fig. 7-3. As described in the text, receptor degradation is treated more accurately with incorporation of ubiquitination. Some details found not to affect qualitative results of interest in this paper were not incorporated in the present study (feedback loops regulating Lck activity (Altan-Bonnet and Germain, 2005; Stefanova et al., 2003) and downstream signaling molecules (Ras, Rac, etc. that are downstream of ZAP70 activation).

cSMAC is modeled as described previously (Lee et al., 2003). We link cSMAC formation to downstream signaling by turning on a potential field that acts on certain species and drives them to form a cSMAC when activated ZAP70 exceeds a threshold. The spatially varying potential field is a coarse-grained representation of forces (derivative of a potential) that direct cSMAC formation. The magnitude of the potential is chosen so that a cSMAC forms in a few (~ 5) minutes. The depth and spatial variation of this potential field determine the rate of cSMAC formation and the concentration of species therein.

The spatio-temporal distribution of each species varies due to diffusion, biochemical reactions that are part of the signaling cascade, and convective motion if a cSMAC is formed. The relevant equations are provided. The numerical methods used to solve the large set of partial differential equations are also described.

Chapter 7.4.1 Signaling reactions used in the simulations

The chemical reactions used to model T cell signal transduction in the computer simulations are detailed below. In our simulations we model the following signaling reactions explicitly: TCR binding and unbinding with pMHC, association and disassociation of Lck, phosphorylation of the TCR ζ chain by bound Lck, association and disassociation of ZAP70 to phosphorylated TCR ζ chains, phosphorylation of ZAP70 by Lck, association and disassociation of ubiquitin ligase Cbl, ubiquitination of activated ZAP70 and proximal TCR ζ chains, and degradation of ubiquitinated species. A generic phosphatase can dephosphorylated activated species. Fig. 7-2 shows a schematic representation of the signaling pathway that was simulated. It is similar to that used by Lee et al. (Lee et al., 2003).

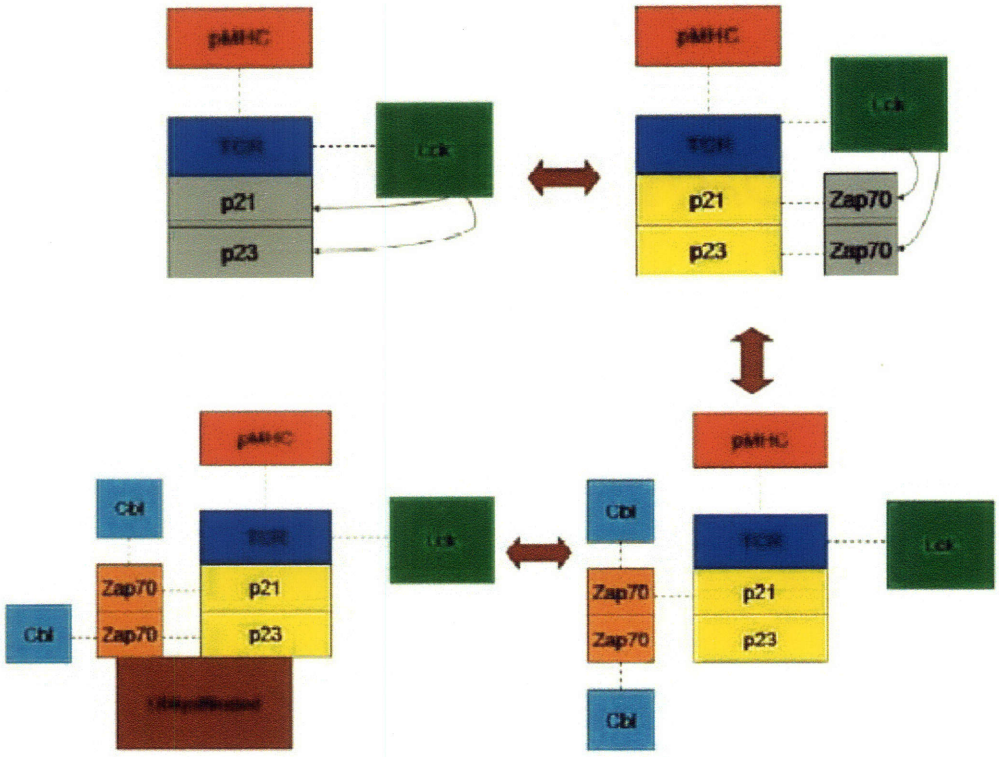


Figure 7-2. **Schematic of the signaling cascade used in the computer simulations**

Dotted lines indicate binding and unbinding events. Arrows denote phosphorylation of inactive species by Lck when it is bound to the signaling complex. Colored boxes represent active states of the proteins. Grey indicates a protein in its inactive state.Brown arrows depict successive events along the signaling cascade. Ultimately, ubiquitinated complexes are degraded. A description of all 324 reaction events used in the simulations is given below.

These events result in twelve distinct classes of half reactions shown below.

KEY:
T TCR
M MHC
L Lck
Z ZAP70
C Cbl

Notes:

1.) Multiple characters in sequence indicate the formation of a complex containing those species attributed to each character.
2.) * following T denotes partial phosphorylation of the TCR ζ chain, * following Z denotes phosphorylaton of SH2 domains of ZAP70, ** following T denotes full phosphorylation of TCR ζ chains.
3.) ()' denotes Ubiquitination of the signaling complex.

TCR- MHC Binding

T + M -> TM
T* + M -> T*M
T** + M -> T**M
T*Z + M -> T*MZ
T**Z + M -> T**MZ
T*Z* + M -> T*MZ*
T**Z* + M -> T**MZ*
T*Z*C + M -> T*MZ*C
T**Z*C + M -> T**MZ*C
T**ZZ + M -> T**MZZ
T**Z*Z + M -> T**MZ*Z
T**ZZ* + M -> T**MZZ*
T**Z*Z* + M -> T**MZ*Z*
T**Z*ZC + M -> T**MZ*ZC
T**ZZ*C + M -> T**MZZ*C
T**Z*Z*C + M -> T**MZ*Z*C
T**Z*Z*CC + M -> T**MZ*Z*CC

TCR- MHC Unbinding

TM -> T + M
T*M -> T* + M
T**M -> T** + M
T*MZ -> T*Z + M
T**MZ -> T**Z + M
T*MZ* -> T*Z* + M
T**MZ* -> T**Z* + M
T*MZ*C -> T*ZZ*C + M
T**MZ*C -> TZZC + M
T**MZZ -> T**ZZ + M
T**MZ*Z -> TZ*Z + M
T**MZZ* -> TZZ* + M
T**MZ**Z** -> TZZ + M
T*MZ*ZC -> T*Z*ZC + M
T*MZZ*C -> T*ZZ*C + M
T*MZ*Z*C -> T*Z*Z*C + M
T**MZ*Z*CC -> T**Z*Z*CC + M

Lck Association (Binding)

TM + L -> TML
T*M + L -> T*ML
T**M + L -> T**ML
T*MZ + L -> T*MLZ
T**MZ + L -> T**MLZ
T*MZ* + L -> T*MLZ*
T**MZ* + L -> T**MLZ*
T*MZ*C + L -> T*MLZ*C
T**MZ*C + L -> T**MLZ*C
T**MZZ + L -> T**MLZZ
T**MZ*Z + L -> T**MLZ*Z
T**MZZ* + L -> T**MLZZ*
T**MZ*Z* + L -> T**MLZ*Z*
T**MZ*ZC + L -> T**MLZ*ZC
T**MZZ*C + L -> T**MLZZ*C
T**MZ*Z*C + L -> T**MLZ*Z*C
T**MZ*Z*CC + L -> T**MLZ*Z*CC

Lck Disassociation (Unbinding)

TML -> TM + L
T*ML -> T*M + L
T**ML -> T**M + L
T*MLZ -> T*MZ + L
T**MLZ -> T**MZ + L
T*MLZ* -> T*MZ* + L
T**MLZ* -> T**MZ* + L
T*MLZ*C -> T*MZ*C + L
T**MLZ*C -> T**MZ*C + L
T**MLZZ -> T**MZZ + L
T**MLZ*Z -> T**MZ*Z + L
T**MLZZ* -> T**MZZ* + L
T**MLZ*Z* -> T**MZ*Z* + L
T**MLZ*ZC -> T**MZ*ZC + L
T**MLZZ*C -> T**MZZ*C + L
T**MLZ*Z*C -> T**MZ*Z*C + L
T**MLZ*Z*CC -> T**MZ*Z*CC + L

ZAP70 Association (Binding)

T* + Z -> T*Z
T* + Z* -> T*Z*
T** + Z -> T**Z
T** + Z* -> T**Z*
T*M + Z -> T*MZ
T*M + Z* -> T*MZ*
T**M + Z -> T**MZ
T**M + Z* -> T**MZ*
T*ML + Z -> T*MLZ
T*ML + Z* -> T*MLZ*
T**ML + Z -> T**MLZ
T**ML + Z* -> T**MLZ*
T**MLZ + Z -> T**MLZZ
T**MLZ + Z* -> T**MLZZ*
T**MLZ + Z* -> T**MLZ*Z
T**MLZ* + Z -> T**MLZ*Z
T**MLZ* + Z -> T**MLZZ*
T**MLZ* + Z* -> T**MLZ*Z*
T**Z + Z -> T**ZZ
T**Z + Z* -> T**Z*Z
T**Z + Z* -> T**ZZ*
T**Z* + Z -> T**Z*Z
T**Z* + Z -> T**ZZ*
T**Z* + Z* -> T**Z*Z*
T**Z*C + Z -> T**Z*ZC
T**Z*C + Z -> T**ZZ*C
T**Z*C + Z* -> T**Z*Z*C
T**MZ + Z -> T**MZZ
T**MZ + Z* -> T**MZ*Z
T**MZ + Z* -> T**MZZ*
T**MZ* + Z -> T**MZ*Z
T**MZ* + Z -> T**MZZ*
T**MZ* + Z* -> T**MZ*Z*
T**MZ*C + Z -> T**MZ*ZC
T**MZ*C + Z -> T**MZZ*C
T**MZ*C + Z* -> T**MZ*Z*C
T**MLZ*C + Z -> T**MLZ*ZC
T**MLZ*C + Z -> T**MLZZ*C
T**MLZ*C + Z* -> T**MLZ*Z*C
(T**MLZ*C)' + Z -> (T**MLZ*ZC)'
(T**MLZ*C)' + Z -> (T**MLZZ*C)'
(T**MLZ*C)' + Z* -> (T**MLZ*Z*C)'
(T**Z*C)' + Z -> (T**Z*ZC)'
(T**Z*C)' + Z -> (T**ZZ*C)'
(T**Z*C)' + Z* -> (T**Z*Z*C)'

(T**MZ*C)˙ + Z -> (T**MZ*ZC)˙
(T**MZ*C)˙ + Z -> (T**MZZ*C)˙
(T**MZ*C)˙ + Z* -> (T**MZ*Z*C)˙

ZAP70 Disassociation (Unbinding)

T*Z -> T* – Z
T*Z* -> T* – Z*
T**Z -> T** + Z
T**Z* -> T** – Z*
T*MZ -> T*M – Z
T*MZ* -> T*M – Z*
T**MZ -> T**M + Z
T**MZ* -> T**M – Z*
T*MLZ -> T*ML – Z
T*MLZ* -> T*ML – Z*
T**MLZ -> T**ML – Z
T**MLZ* -> T**ML – Z*
T**MLZZ -> T**MLZ – Z
T**MLZZ* -> T**MLZ + Z*
T**MLZ*Z -> T**MLZ* – Z
T**MLZ*Z -> T**MLZ + Z*
T**MLZZ* -> T**MLZ* – Z
T**MLZ*Z* -> T**MLZ* – Z*
T**ZZ -> T**Z – Z
T**ZZ* -> T**Z + Z*
T**Z*Z -> T**Z* – Z
T**Z*Z -> T**Z + Z*
T**ZZ* -> T**Z* – Z
T**Z*Z* -> T**Z* – Z*
T**Z*ZC -> T**Z*C – Z
T**ZZ*C -> T**Z*C – Z
T**Z*Z*C -> T**Z*C – Z*
T**MZZ -> T**MZ – Z
T**MZ*Z -> T**MZ – Z*
T**MZZ* -> T**MZ – Z*
T**MZ*Z -> T**MZ* – Z
T**MZZ* -> T**MZ* – Z
T**MZ*Z* -> T**MZ* – Z*
T**MZ*ZC -> T**MZ*C – Z
T**MZZ*C -> T**MZ*C – Z
T**MZ*Z*C -> T**MZ*C – Z*
T**MLZ*ZC -> T**MLZ*C + Z
T**MLZZ*C -> T**MLZ*C + Z
T**MLZ*Z*C -> T**MLZ*C – Z*
(T**MLZ*ZC)˙ -> (T**MLZ*C)˙ – Z
(T**MLZZ*C)˙ -> (T**MLZ*C)˙ – Z
(T**MLZ*Z*C)˙ -> (T**MLZ*C)˙ + Z*
(T**Z*ZC)˙ -> (T**Z*C)˙ – Z
(T**ZZ*C)˙ -> (T**Z*C)˙ + Z
(T**Z*Z*C)˙ -> (T**Z*C)˙ + Z*
(T**MZ*ZC)˙ -> (T**MZ*C)˙ – Z
(T**MZZ*C)˙ -> (T**MZ*C)˙ – Z
(T**MZ*Z*C)˙ -> (T**MZ*C)˙ + Z*

Cbl Association (Binding)

T*Z* + C -> T*ZC
T**Z* + C -> T**ZC
T**Z*Z + C -> T**Z*ZC
T**ZZ* + C -> T**ZZ*C
T**Z*Z* + C -> T**Z*Z*C
T**Z*Z*C + C -> T**Z*Z*CC
T*MZ* + C -> T*MZ*C
T**MZ* + C -> T**MZ*C
T**MZ*Z + C -> T**MZ*ZC
T**MZZ* + C -> T**MZZ*C
T**MZ*Z* + C -> T**MZ*Z*C
T**MZ*Z*C + C -> T**MZ*Z*CC
T*MLZ* + C -> T*MLZ*C
T**MLZ* + C -> T**MLZ*C
T**MLZ*Z + C -> T**MLZ*ZC
T**MLZZ* + C -> T**MLZZ*C
T**MLZ*Z* + C -> T**MLZ*Z*C
T**MLZ*Z*C + C -> T**MLZ*Z*CC

Cbl Disassociation (Unbinding)

T*Z*C -> T*Z* + C
T**Z*C -> T**Z* + C
T**Z*ZC -> T**Z*Z + C
T**ZZ*C -> T**ZZ* + C
T**Z*Z*C -> T**Z*Z* + C
T**Z*Z*CC -> T**Z*Z*C + C
T*MZ*C -> T*MZ* + C
T**MZ*C -> T**MZ* + C
T**MZ*ZC -> T**MZ*Z + C
T**MZZ*C -> T**MZZ* + C
T**MZ*Z*C -> T**MZ*Z* + C
T**MZ*Z*CC -> T**MZ*Z*C + C
T*MLZ*C -> T*MLZ* + C
T**MLZ*C -> T**MLZ* + C
T**MLZ*ZC -> TMLZ*Z + C
T**MLZZ*C -> TMLZZ* + C
T**MLZ*Z*C -> TMLZ*Z* + C
T**MLZ*Z*CC-> TMLZ*Z*C + C

Phosphorylation of TCR ζ Chains and ZAP70 SH3 domains by Lck

TML -> T*ML
T*ML -> T**ML
T*MLZ - > T**MLZ
T*MLZ -> T*MLZ*
T**MLZ -> T**MLZ*
T**MLZZ -> T**MLZ*Z
T**MLZZ -> T**MLZZ*
T**MLZ*Z -> T**MLZ*Z*
T**MLZZ* -> T**MLZ*Z*
T*MLZ*C -> T**MLZ*C
T**MLZZC -> T**MLZ*ZC
T**MLZZC -> T**MLZZ*C
T**MLZ*ZC -> T**MLZ*Z*C
T**MLZZ*C -> T**MLZ*Z*C

Dephosphorylation by Phosphatases

T* -> T
T** -> T*
Z* -> Z
T*M -> TM
T**M -> T*M
T*Z* -> T*Z
T**Z* -> T**Z
T**Z -> T*Z
T**Z* -> T*Z*
T**Z*Z -> T**ZZ
T**ZZ* -> T**ZZ
T**Z*Z* -> T**Z*Z
T**Z*Z* -> T**ZZ*
T*MZ* -> T*MZ
T**MZ* -> T**MZ
T**MZ -> T*MZ
T**MZ* -> T*MZ*
T**MZ*Z -> T**MZZ
T**MZZ* -> T**MZZ
T**MZ*Z* -> T**MZ*Z
T**MZ*Z* -> T**MZZ*
T**MLZ* -> T**MLZ
T**MLZ -> T*MLZ
T**MLZ* -> T*MLZ*
T**MLZ*Z -> T**MLZZ
T**MLZZ* -> T**MLZZ
T**MLZ*Z* -> T**MLZ*Z
T**MLZ*Z* -> T**MLZZ*
T**MLZ*Z*C -> T**MLZ*ZC
T**MLZ*Z*C -> T**MLZZ*C

Ubiquitination of Signaling Complex by Cbl

T*Z*C -> (T*Z*C)'
T**Z*C -> (T**Z*C)'
T**Z*ZC -> (T**Z*ZC)'
T**ZZ*C -> (T**ZZ*C)'
T**Z*Z*C -> (T**Z*Z*C)'
T**Z*Z*CC -> (T**Z*Z*CC)'
T*MZ*C -> (T*MZ*C)'
T**MZ*C -> (T**MZ*C)'
T**MZ*ZC -> (T**MZ*ZC)'
T**MZZ*C -> (T**MZZ*C)'
T**MZ*Z*C -> (T**MZ*Z*C)'
T**MZ*ZC -> (T**MZ*ZC)'
T**MZZ*C -> (T**MZZ*C)'
T**MZ*Z*C -> (T**MZ*Z*C)'
T**MZ*Z*CC -> (T**MZ*Z*CC)'
T*MLZ*C -> (T*MLZ*C)'
T**MLZ*C -> (T**MLZ*C)'
T**MLZ*ZC -> (T**MLZ*ZC)'
T**MLZZ*C -> (T**MLZZ*C)'
T**MLZ*Z*C -> (T**MLZ*Z*C)'
T**MLZ*Z*CC -> (T**MLZ*Z*CC)'

Degradation of Ubiquitinated Species

(T*Z*C)' -> C
(T**Z*C)' -> C
(T**Z*ZC)' -> C
(T**ZZ*C)' -> C
(T**Z*Z*C)' -> C
(T**Z*Z*CC)' -> 2C
(T*MZ*C)' -> M + C
(T**MZ*C)' -> M + C
(T**MZ*ZC)' -> M + C
(T**MZZ*C)' -> M + C
(T**MZ*Z*C)' -> M + C
(T**MZ*Z*CC)' -> M + 2C
(T*MLZ*C)' -> M + C + L
(T**MLZ*C)' -> M + C + L
(T**MLZ*ZC)' -> M + C + L
(T**MLZZ*C)' -> M + C + L
(T**MLZ*Z*C)' -> M + C + L
(T**MLZ*Z*CC)' -> M + 2C + L

Chapter 7.4.2 Numerical solutions of the equations

The Smoluchowski equations that describe the system can be illustrated by considering the following reaction: A + B -> C. For instance, A could be TCR, B could be pMHC, and C could be the TCR-pMHC complex. The equation governing any one of the reactant species (say, A) is:

$$\frac{\partial \rho_i}{\partial t} = D_i \nabla^2 \rho_i - k_{on} \rho_i \rho_j + k_{off} \rho_k - a_i \Theta(\rho_d - \rho_d^*) \nabla \cdot [(\nabla U) \rho_i] \quad (7\text{-}1)$$

where ρ_i and ρ_j are the densities of reactants (like A and B) and ρ_k is the density of a product. k_{on} and k_{off} are the rate parameters corresponding to association and dissociation, respectively. The signs preceding the reaction kinetic terms would be the opposite if we were writing an equation for the density of a product species. If the species, i, is subjected to directed motion (e.g., TCR), then a_i equals unity, otherwise it is zero. θ is a step function that equals unity if the density of a downstream signaling product, ρ_d, exceeds a threshold value (ρ_{d^*}). U is the spatially varying potential that drives clustering and D_i is the diffusion coefficient of the species, i.

We use Euler discretization in space and time to solve the Smoluchowski type equations. The density fields are discretized on a three dimensional cubic grid, with lattice spacings, Δx, Δy, and Δz, along x, y and z axes, respectively. We choose periodic boundary conditions in the x and y directions and constant density boundary condition in the z directions. These choices are consistent with the physical situation, where the z=0 plane represents junction between the T-cell surface and the

Antigen presenting cell (APC). We have taken $\Delta x = \Delta y = \Delta z = 2$ and $\Delta t = 0.01$, which gives us a stable solution, and our results do not depend on the small variations around values of Δx, Δy, Δz and Δt.

Chapter 8 – Signaling Networks at a Global Scale: Analysis of Phosphoproteomics Data Obtained from Mass Spectrometry

Chapter 8.1. Introduction

A detailed understanding of the overarching principles that underlie how mammalian cells detect, integrate, and utilize external signals to achieve an appropriate phenotypic has been a longstanding challenge in the era of post-genomic biology. Signals derived from activated receptor tyrosine kinases lead to the phosphorylation of hundreds of tyrosine residues that are embedded within complex biochemical signaling networks. Recent technological advances in the application of mass spectrometry in the emerging field of phosphoproteomics have allowed for the quantitative measurement, in real time, of the dynamics of the activity of hundreds of tyrosine residues as they undergo reversible phosphorylation in response to a stimulus such as a growth factor[165-168]. These measurements indirectly probe the structure and dynamics of these signaling networks, in which the phosphorylation sites are embedded.

The enormous amount of quantitative data acquired in these experiments raises the question of what modeling approaches might be used to lend predictive and mechanistic insight into the signaling networks that govern the dynamics of these phosphorylated tyrosine (pY) sites[169]. Clustering and similarity and other measures of correlation have successfully grouped large data sets including data derived from mass spec-measured signaling dynamics into similar patterns. Partial least squares regression modeling (PLSR) has also been employed to identify what aspects of these data sets are most correlated with different phenotypic responses. These statistical techniques have shown to be very useful in their predictive capabilities and have yielded new biological insights[132].

In spite of their predictive power, causality and other prerequisites for mechanistic understanding are often lost in the interpretation of measures of correlation. In an alternative approach, we develop and apply techniques rooted in statistical mechanics and information theory to investigate the nature of both the global features and local interactions of these signaling networks. In developing these techniques, we hope to first introduce new frameworks for understanding, mechanistically, how these networks integrate signals and how deregulation within these networks leads to myriad diseases such as cancer and autoimmune disorders, we then hope to exploit these new frameworks to generate novel biological hypotheses. One approach that we use utilizes the principle of maximum entropy to infer an interaction network among the phosphotyrosine (pY) sites.

Models of pairwise network connectivity obtained from entropy maximization have proven insightful in different and seemingly disparate contexts[170, 171]. In one example, Bialek and coworkers applied the principle of entropy maximization to construct an interaction network of neurons that respond to visual stimuli[171]. In another example, Fedoroff and coworkers use the principle to derive a genetic interaction network from microarray data in Yeast. In this study, we use the principle to investigate the network connectivity of phosphotyrosine interaction networks.

At the local scale, we seek to infer the network of interacting tyrosine sites that governs signaling dynamics. We construct the interaction network between the phosphorylated sites by developing a model of network connectivity that maximizes the Shannon entropy[11] subject to the constraints of the pairwise correlations present in the data set. In developing this model, we predict the existence of few phosphotyrosine

"hubs" that interact with many other tyrosine sites. The model also can be used to characterize the affinity or interaction strength dependent topology of the phosphotyrosine signaling networks. For intermediate values of an interaction strength threshold, a small world network topology is observed. Finally, the model makes predictions of network connectivity between novel phosphotryosine sites identified by mass spectrometry and those of known nodes in the cognate signaling network.

Ongoing work in understanding signaling dynamics at the global scale of the entire network is discussed. We first use a singular value decomposition to identity the slow characteristic modes of the signaling network that reveal the gross features of the underlying biochemical network. We then propose speculative work in the development of projection operator methods that aim to collapse all of the data onto these characteristic signaling modes by deriving generalized Langevin equations for describing the dynamics of these collective variables[172-174]. In the future, these equations may serve as a framework for the predicting how these networks respond to perturbations.

Chapter 8.2. Methods

The approach that is taken to derive the network connectivity has its roots in statistical mechanics and information theory. Since there are many more observed phosphorylation sites than independent samples of the network configuration, the problem of finding the unique network connectivity is ill-posed. Instead, the question that is asked is how to calculate the probability of a network configuration (i.e. phosphorylation state of the N Tyrosine sites that are measured). Conceptually, this notion has been used in the context of other complex molecular systems by applying the principles of statistical mechanics and the second law of thermodynamics[36] – i.e. the

macroscopic state (network connectivity) of interest is one that allows for the most number of microscopic network configurations.

In this statistical approach, the microscopic state of the complex system is either not useful or can not be effectively resolved. Rather than resolving the state of the system deterministically, a statistical ensemble in which the state of the system fluctuates and is considered. Therefore, each observation of the state of the system constitutes a sampling from this ensemble. Our maximum entropy approach implies that the probability distribution associated with this ensemble then is the distribution that allows for the greatest number of accessible microscopic states. The number of accessible states or entropy has its analogies in information theory. In information theory, a type of entropy, known as the Shannon entropy[11], is a measure of the amount of uncertainty in a random variable—the greater the number of accessible microscopic states, the greater the uncertainty about the state of the system (the larger the entropy). Shannon entropy is seen as a measure of the number of measurements needed to have a full knowledge of a random variable. Thus, when in this context, assuming the least amount of knowledge of a random variable implies that the entropy of the system is at a global maximum.

From the mass spec data[168], times courses of the relative amount of phosphorylation at each of the $N = 222$ phosphorylated tyrosine sites is obtained at $T = 7$ time points at times $\{0,1,2,4,8,16,32\}$ minutes.

The main assumption in developing our model is that the relative amounts of Tyrosine phosphorylation measured at each pY site and at time point constitute one sample from a statistical ensemble of possible phosphorylation states whose activities fluctuate on an interconnected network. Since time courses are measured up to a time of

approximately one half hour, the differences in phosphorylation levels measured at each time point arise from changes in the amount of reversible post translational modifications. Changes in gene expression occur on longer time scales[10] so we do not expect the time courses to be affected by gene transcription upon which the upregulation of genes will affect the network topology. Therefore, it seems reasonable to expect the same protein-protein interaction network to be present across a time scale of 30 minutes.

We characterize the state of the system with a state vector $\vec{x}$ that contains the relative amount of pY phosphorylation at each measured site;

$$\vec{x} \equiv \{x_i\} = \{x_1, x_1, ..., x_N\}, \tag{8-1}$$

x_i is the amount of Tyrosine phosphorylation at the i^{th} pY site.

A probability of observing the network in a particular configuration $p(\{x_i\})$ is then defined. The entropy H is then defined,

$$H = -\sum_{\{x_i\}} p(\{x_i\}) \ln\left(p(\{x_i\})\right). \tag{8-2}$$

Since it is apparent that the magnitude of fluctuations about the average of each time can greatly differ, data are rescaled to unit variance to focus on the relative shapes of the time courses. Therefore, a new set of scaled variables is considered:

$$\hat{x}_i^k = \frac{x_i^k - \langle x_i \rangle}{\sqrt{\langle x_i x_i \rangle}}, \tag{8-3}$$

where brackets denote an average over samples,

$$\langle x_i \rangle = T^{-1} \sum_{k=1}^{T} x_i^k \tag{8-4}$$

and

$$\left\langle \hat{x}_i \hat{x}_j \right\rangle = T^{-1} \sum_{k=1}^{T} \left(x_i^k - \left\langle x_i \right\rangle \right) \left(x_j^k - \left\langle x_j \right\rangle \right) \tag{8-5}$$

and the superscript k runs from 1 to 7 and denotes the k^{th} time point or measurement. When written in this form, $\left\langle \hat{x}_i \hat{x}_j \right\rangle$ is the Pearson covariance matrix. The new set of random variables is now: $\vec{\hat{x}}_i \equiv \left\{ \hat{x}_i \right\} = \left\{ \hat{x}_1, \hat{x}_2, ..., \hat{x}_N \right\}$. Since our interest is in network connectivity, we consider two point (i.e. pairwise) interactions and therefore the mean and covariance of the data. The task is to maximize the entropy H subject to the constraints:

$$\sum_{\{\hat{x}_i\}} p\left(\left\{ \hat{x}_i \right\} \right) = 1, \tag{8-6}$$

and

$$\left\langle \hat{x}_i \right\rangle = 0, \tag{8-7}$$

and

$$\left\langle \hat{x}_i \hat{x}_j \right\rangle = \frac{\left\langle x_i x_j \right\rangle}{\sqrt{\left\langle x_i x_i \right\rangle \left\langle x_j x_j \right\rangle}} \equiv C_{ij}. \tag{8-8}$$

For each element of the covariance matrix C_{ij}, there is a corresponding Lagrange multiplier J_{ij}; an additional Lagrange multiplier α is required for the normalization condition. The functional $L\left[p\left(\left\{ \hat{x}_i \right\} \right) \right]$ is maximized:

$$L\left[p\left(\left\{ \hat{x}_i \right\} \right) \right] = H\left[p\left(\left\{ \hat{x}_i \right\} \right) \right] - \alpha \sum_{\{\hat{x}_i\}} p\left(\left\{ \hat{x}_i \right\} \right) - \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{2} J_{ij} \sum_{\{\hat{x}_i\}} p\left(\left\{ \hat{x}_i \right\} \right) \hat{x}_i \hat{x}_j \tag{8-9}$$

and

$$\frac{\delta L}{\delta p(\{x_i\})} = 0 \tag{8-10}$$

This procedure results in the following form for $p\left(\{\hat{x}_i\}\right)$:

$$p\left(\{\hat{x}_i\}\right) = Z^{-1} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{x}_i J_{ij}\hat{x}_j\right], \tag{8-11}$$

where Z is a normalization factor. The partition function Z can easily be obtained using standard methods but is not necessary for the purpose of this study. In future work, it may be interesting to study thermodynamic properties of the model by studying the behavior of Z and its logarithm that constitutes a free energy. Also, higher order networks can be obtained by constraining the entropy to higher order moments of the probability distribution[170]. Since the form of the resulting distribution $p\left(\{\hat{x}_i\}\right)$ is a multivariate Gaussian whose argument (energy function) is a summation over all pairwise and self interactions, $p\left(\{\hat{x}_i\}\right)$ can also be written in terms of the elements of the covariance matrix C_{ij}:

$$p\left(\{\hat{x}_i\}\right) = Z^{-1} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{x}_i C_{ij}^{-1}\hat{x}_j\right]. \tag{8-12}$$

The interaction matrix J_{ij} gives the pairwise coupling between each phosphorylation site in the network and is the subject of the analysis. J_{ij} is the inverse of the covariance matrix C_{ij}; i.e. $J_{ij}^{-1} = C_{ij}$.

From the original data, it is noted that there are many more phosphorylation sites N is much larger than the number of time points sampled T ($N > T$); C_{ij} is therefore underdetermined (i.e. the rank of C is not full). Therefore, J_{ij} is obtained by inverting

C_{ij} is the space of non-zero eigenvalues. First, C_{ij} is expanded in terms of its eigenvalues and coefficients of its eigenvectors:

$$C_{ij} = \sum_{k=1}^{N} \omega_i^k \upsilon_i^* \upsilon_j \tag{8-13}$$

where the * denotes the transpose of the eigenvector υ. The matrix element J_{ij} is obtained by inverting C in the space of eigenvectors containing non-zero eigenvalues by considering only the non-zero eigenvalues of C_{ij}:

$$J_{ij} = \sum_{k=1}^{r} \left(\omega_i^k\right)^{-1} \upsilon_i^* \upsilon_j \,. \tag{8-14}$$

r is the rank of the covariance matrix and is equal to $T-1; r = T-1$. From the expansions in eqs. 8-13 and 8-14, it is clear that the correlations are dominated by the largest eigenvalues whereas the interactions are dominated by the smallest eigenvalues. The matrix elements J_{ij} are interpreted as the residual, pairwise interactions between phosphotyrosine (pY) sites i and j that remain once global effects that are dominated by the correlations (i.e. the largest eigenvalues) are removed. J_{ij} constitutes the matrix of pairwise network interactions. A schematic of this effect is presented in Fig. 8-1. In Fig. 8-1a, two signals S_1 and S_2 act on pY site i and a third signal S_3 acts on site j. Also, site j acts on site i. In this scenario, the interaction at sites i and j is not necessarily revealed by the correlation since other factors are influencing the correlation. Those influences are present in the larger eigenvalues of the covariance matrix C and are removed in the computation of the interaction matrix J. In Fig. 8-1b, a signal S acts on both pY sites i and j. In this scenario, i and j have a small (nonexistent) interaction. However the activity at sites i and j can be highly correlated since they are activated by the same

signal. This effect is contained in the large eigenvalues of C. Fig. 8-1c contains a scatter plot of the elements of J plotted against the elements of C. A nontrivial relationship between the elements of J and C is observed.

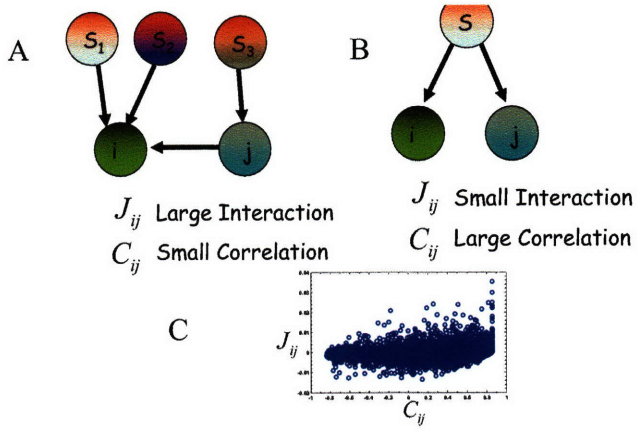


Figure 8-1. **Illustration of the differences between the elements of the correlation and interaction matrix.**

a.) an example in which pY sites i and j can have a large interaction but small correlation. b.) an example in which pY sites i and j can have a small (nonexistent) interaction but large correlation. c.) a scatterplot of the matrix elements of J plotted against the matrix elements of C.

Chapter 8.2.1 Validation of Maximum Entropy model for network connectivity

It is important to note that the measured time courses from the mass spec data are not stationary and therefore it is not entirely clear why a maximum entropy based model constrained to pairwise correlations in a time series should be should be applicable.

We first tested the applicability of this approach to reconstructing network connectivity by simulating signaling dynamics on a model signaling network with known topology. A simplified model of a signaling cascade was considered and a ordinary differential equation model subject to the topology of the model signaling network was

developed. The model consists of a kinase cascade with branch points and multiple layers of positive and negative feedback control. Kinetic equations are modeled according to the weakly activated cascade approximation which is seen as a limiting case of Michaelis Menten kinetics[31]. Data will be reported in a future manuscript. Future work also will consider the simulation of several standard signaling models and then sampling from their simulated outputs to reconstruct the prescribed network topology. From this procedure, a systematic evaluation of the successes and failures of this approach can be obtained.

Chapter 8.3. Results

Chapter 8.3.1 Affinity dependent network topology

The interaction matrix J_{ij} encodes a total of $\binom{N}{2} = \frac{N(N-1)}{2}$ pairwise interactions among the phosphotyrosine sites. For $N = 222$ sites, 24,531 interactions are possible. As shown in Fig. 8-1, the probability distribution $P(J_{ij})$ is sharply peeked at zero.

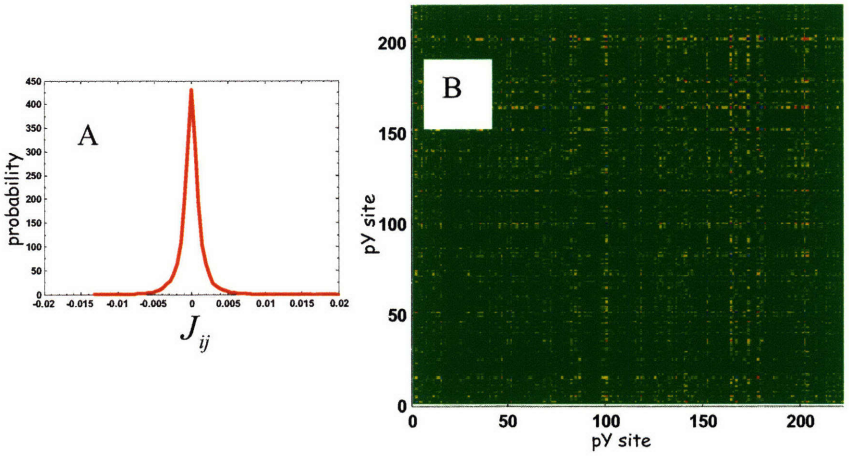


Figure 8-2. **Pairwise network interactions**

a.) $P(J_{ij})$ is plotted as a function of J_{ij}. $P(J_{ij})$ is sharply peeked at zero as seen in its rapid decay away from $J_{ij} = 0$ b.) Contour plot of the interaction matrix J.

To define a network topology, a threshold parameter T_k is introduced. For a given value of T_k, an interaction between pY sites at positions i and j is counted if the magnitude of J_{ij} exceeds the threshold; $|J_{ij}| > T_k$. A connectivity matrix

$$K_{ij} \equiv J_{ij}\Theta(|J_{ij}| - T_k) \qquad (8\text{-}15)$$

, where $\Theta(X)$ is a Heaviside step function, is used to define the network connectivity. The resulting network, therefore, is highly dependent on the choice of threshold T_k. Fig. 8-2 illustrates the dependence of the network topology on choice of T_k. In Fig. 8-2, the node distribution $P(k)$ is plotted for four different values of T_k. $P(k)$.is defined as the probability that a pY site is connected to k separate sites. At high thresholds (e.g. $T_1 = 10.0x10^{-3}$) most signaling nodes have no connections. At low thresholds (e.g. $T_4 = 2.5x10^{-2}$), each phosphotyrosine site is connected to many other sites. At intermediate values of T_k (e.g. $T_2 = 5.0x10^{-2}$ or $T_3 = 7.5x10^{-2}$), $P(k)$ appears as a monotonically decaying, continuous function of k. Unfortunately, due to the insufficient number of samples of k, the large k behavior of $P(k)$ is not resolved. It is apparent, however, that at these intermediate threshold values, the presence of a significant tail of the distribution $P(k)$ is apparent. Also at intermediated threshold values, $P(k)$ is seen as a monotonically decaying function k.

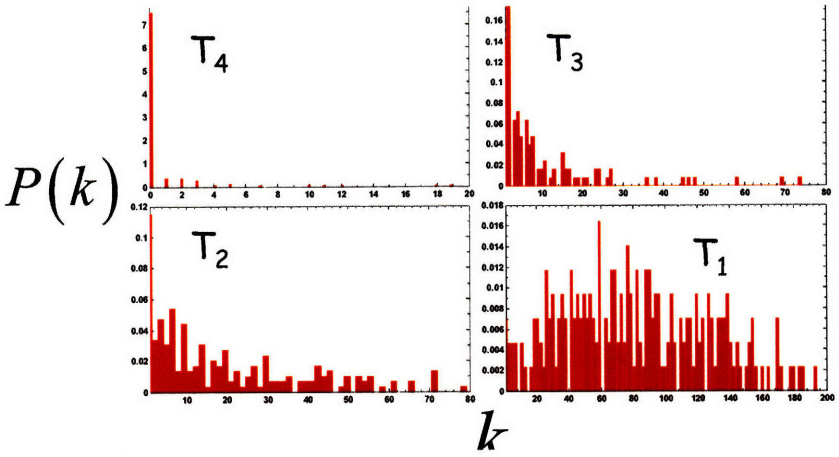


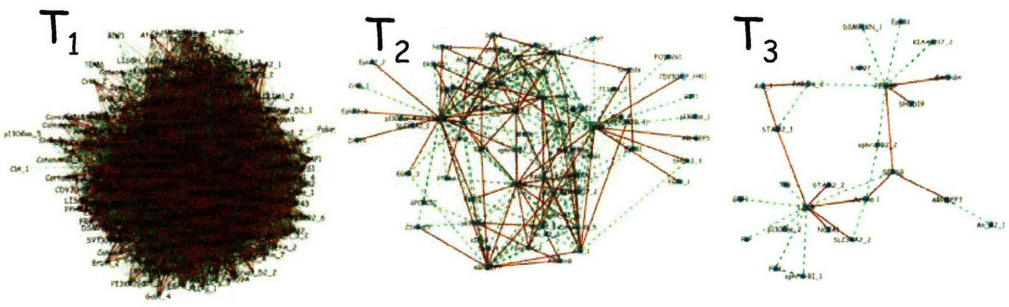
Figure 8-3. Threshold dependence of network topology

The node distribution (i.e. probability that a phosphorylation site is linked to k other sites) $P(k)$ is plotted for four values of the threshold T_k; an interaction between sites i and j is counted if $|J_{ij}| > T_k$ a.) $T_1 = 2.5x10^{-3}$, b.) $T_2 = 5.0x10^{-2}$, c.) $T_3 = 7.5x10^{-2}$, d.) $T_4 = 10.0x10^{-2}$.

Chapter 8.3.2 Maximum Entropy model predicts a core network structure

Fig 8-4. considers graphical depictions of the pY-pY phosphotyrosine interaction network at different thresholds. At a low threshold, $T_1 = 2.5x10^{-3}$, the network is highly interconnected and not easily interpretable. At a high threshold, $T_3 = 7.5x10^{-2}$, a small set of interconnected nodes is present. The nodes consist of a set phosphotyrosine sites that are connected together in a 'small-world' fashion in the sense that each pY is closely

connected to each other site[175]. This behavior would not be expected from a graph of a randomly connected network with as many connections. Instead, it is apparent that the mean path length between two nodes of the network is much shorter than that of a random network. Such 'small-world' behavior is also exhibited at intermediate in which more connectivity among the signaling nodes is resolved. A closer inspection of the network at intermediate threshold (Fig. 8-4d.) reveals interconnected phosphotyrosine sites of widely disparate functional significance. The core structure of the signaling network contains pY sites on proteins involved, in endocytosis, gene splicing, MAPK signaling, PI3K/AKT signaling, among others.



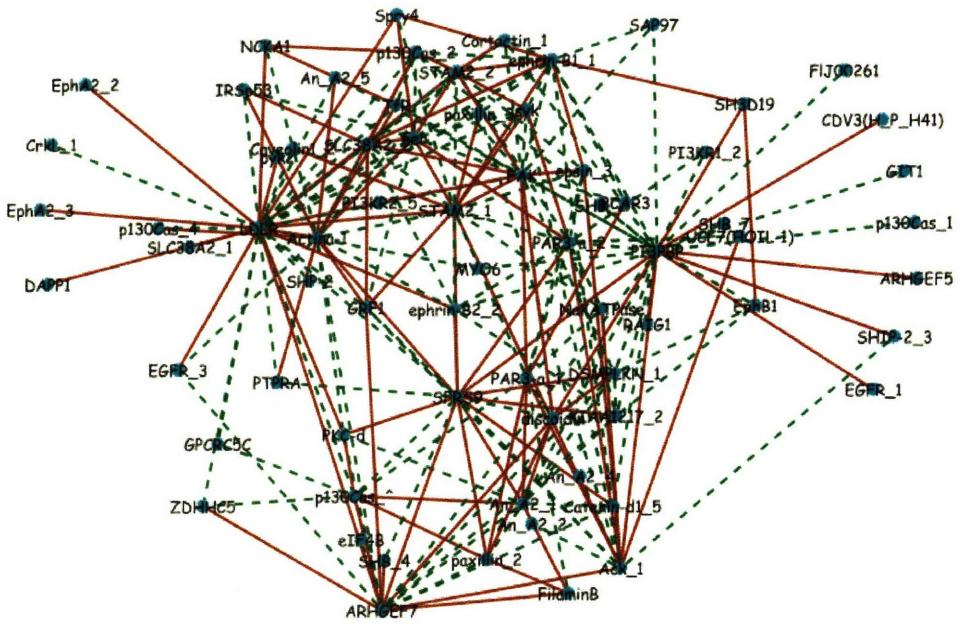


Figure 8-4. Network connectivity at different thresholds

A graphical display of the pairwise pY-pY interaction network. Dotted green lines denote positive interactions. Solid red lines denote negative interactions. Three values of the threshold are shown a.) $T_1 = 2.5x10^{-2}$ b.) $T_2 = 5.0x10^{-2}$ and c.) $T_3 = 7.5x10^{-2}$. d.) A closeup view of the network at intermediate threshold T_2.

Furthermore, an inspection of the diagonal elements of the interaction matrix contains information about the 'self' interactions, J_{ii}. The elements J_{ii} are a measure of the overall contribution of the ith pY site to the structure of the network. The sites with the largest self interaction can be considered the network hubs. Table 8-1 contains a list of the 10 largest self interactions. These network 'hubs' are well represented in the core structure of the signaling network in Fig. 8-4d. Also note that since the data are scaled to unit variance, pY sites with small variance with contribute more to the network structure. Therefore, sites with small variance (on the order of the error bars of the experiment) are not considered. For example, the pY site of GSK3-β is constitutively active and not

considered. The inverse of the variance at each site is also shown in Table 8-1.

Name	Self Interaction ($x10^{-2}$)	1/Variance ($x10^{2}$)
PI3BP_Y492	2.51	7.06
LDLR	1.55	2.14
ARHGEF7	1.14	7.33
PAR3aY1127	1.11	2.02
PAR3aY1080	1.09	9.61
FAK	1.03	1.06
Actin-a1	1.03	5.88
ACK_Y857/Y858	1.20	0.06
SFRS9	1.01	4.98
STAM2_Y292	0.94	0.03

Table 8-1. **List of the 10 pY sites with largest self interaction and sufficiently large variance.**

Chapter 8.3.3 Analysis of previously uncharacterized signaling nodes

Of the N=222 detected phosphorylated peptides that comprise the phosphotyrosine signaling network, 31 have been previously unassociated with the EGFR signaling network[168]. The network connectivity also makes predictions about the connectivity of the uncharacterized nodes. Table 8-2. lists the 5 largest self interactions among the pY sites that have not been associated with the network. These are the five sites that are most influential to forming the structure of the network as determined by their values of J_{ii}

Name	Self Interaction ($x10^{-2}$)	1/Variance ($x10^{2}$)
ARHGEF7	1.14	7.33
SH3D19	0.70	1.26
ZDHHC5	0.55	2.87
An A2	0.49	2.73
GPCRC5C	0.45	2.77

Table 8-2. Five most strongly interacting uncharacterized phosphotyrosine sites

The highest scoring uncharacterized pY site is associated with a protein ARHGEF7. ARHGEF7 is a guanine exchange factor associated with the Rho GTPase Rac and its pathway[176]. The Rac pathway is involved in cytoskeletal rearrangement and cell motility among many other functions. ARHGEF7 is a putative tumor suppressor protein that is associated with the mammalian scribble protein. The next highest scoring pY site is associated with a poorly characterized SH3 domain containing protein SH3D19. SH3D19 is a likely adapter protein that contains 5 SH3 domains[177]. The third highest scoring pY site is associated with a zinc finger domain containing protein ZDHHC5. ZDHHC5 was identified in a shRNA screen for genes involved in cell division in HeLa cells[178]. The next highest scoring pY site is associated with a protein known as An A2. An A2 is a phospholipase family and has numerous functions downstream of the generation of lipid secondary messengers[179]. Finally, the fifth highest scoring pY site is associated with GPCRC5C. GPCR5C is a tissue specific, G-protein coupled receptor[180]. Fig. 8-5 contains a graphical representation of the network connectivity of the three highest scoring pY sites, those associated with:

ARHGEF7, SH3D19, and ZDHHC5.

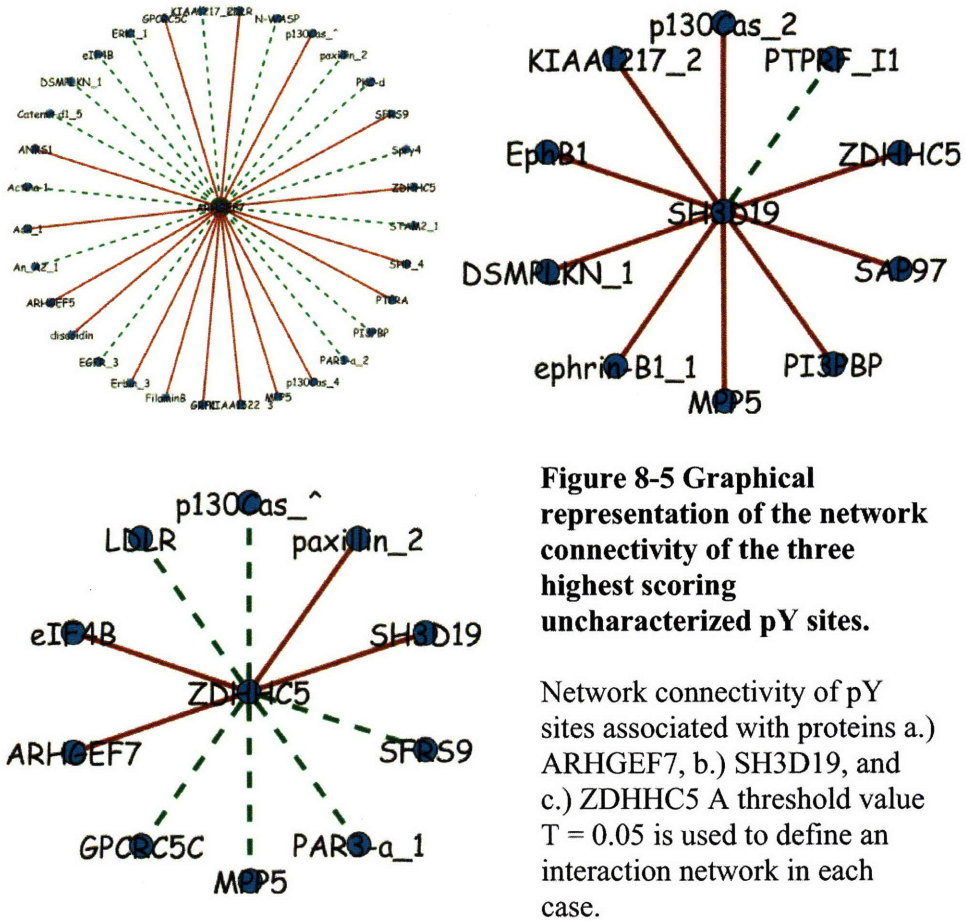


Figure 8-5 Graphical representation of the network connectivity of the three highest scoring uncharacterized pY sites.

Network connectivity of pY sites associated with proteins a.) ARHGEF7, b.) SH3D19, and c.) ZDHHC5 A threshold value T = 0.05 is used to define an interaction network in each case.

Chapter 8.3.4 A singular value decomposition reveals the presence of slow dynamic modes in the data

In the work discussed above, we sought to reconstruct from entropic considerations an interaction network of the phosphotyrosine sites. The goal of such an analysis was to characterize the short length scale behavior of the phosphotyrosine interaction network. In this section, we discuss work in progress that aims to characterize the global (large length scale) behavior of the signaling network.

We first compute the singular value decomposition of the N x T matrix A whose rows constitute the N pY sites and whose columns comprise the T total time points.

$$A = U\Lambda\upsilon^T \equiv UX \tag{8-16}$$

where,

$$\left(A^T A\right)\upsilon = \lambda\upsilon \tag{8-17}$$

and

$$A\upsilon = \lambda U\,. \tag{8-18}$$

So that:

$$A_{ij} = \sum_k U_{ik} X_{kj}\,. \tag{8-19}$$

Or written more suggestively,

$$A_i(t) = \sum_k U_{ik} X_k(t)\,. \tag{8-20}$$

Fig. 8-6a contains a ranking of the singular values contained in the first T diagonal elements of the Λ matrix. Each eigenvalues is rescaled to the sum of the T total eigenvalues. When viewed in this manner, it is apparent that the two largest modes constitute over 75% of the sum total of all eigenvalues. Therefore, we focus on the first two modes. Fig. 8-6b contains a plot of the first two (k = 1 and k = 2) modes $X_1(t)$ and $X_2(t)$. A

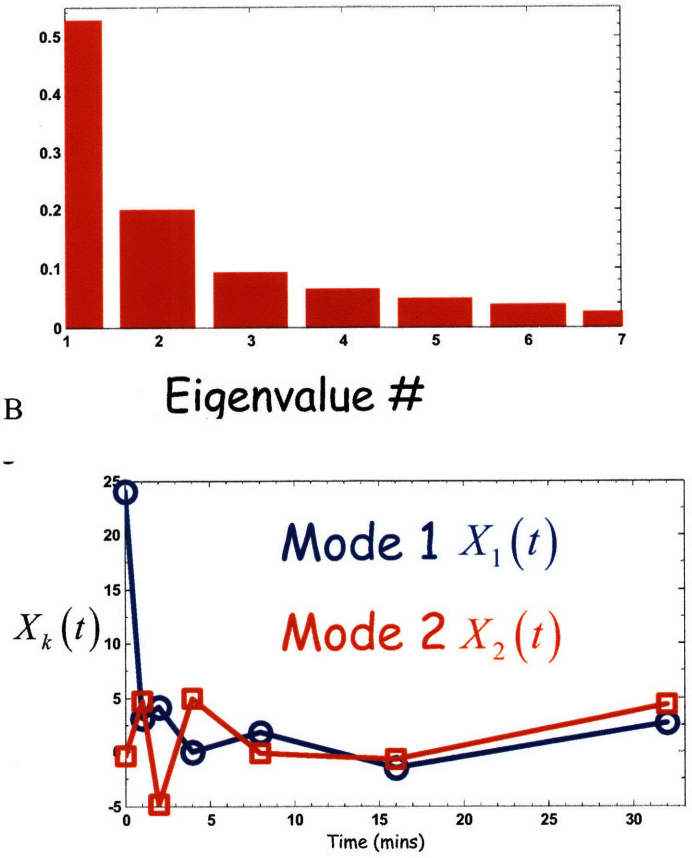


Figure 8-6. **Singular value decomposition of the pY data.**

a.) A plot of the T eigenvalues. Each eigenvalues is scaled to the sum of the T eigenvalues. b.) a plot of the first two (k = 1 and k = 2) modes $X_1(t)$ and $X_2(t)$.

From the plot in Fig. 8-6b, we observe that the first mode decays over time to what may be a steady state value. The second mode first oscillates and then apparently reaches a similar plateau. Further work will be required to physically interpret the meaning of these two modes. In making such an interpretation, we introduce a new description of how to characterize signaling dynamics. One potential way of interpreting the behavior in Fig 8-6. would be to first fit a set of differential equations to the set of data; i.e.

$$\frac{d}{dt}A_i = F_i\left(\{A_1, A_2, ..., A_N, t\}\right) \tag{8-21}$$

where $A_i(t)$ is the activity of the ith phosphorylation site. Then projection operator methods can be applied to the N = 222 equations that constitute the signaling dynamics to project these dynamics onto the two characteristic modes $X_1(t)$ and $X_2(t)$:

$$\{A_i(t)\} \rightarrow \{X_1(t), X_2(t)\}. \tag{8-22}$$

The resulting generalized Langevin equations may be used to describe the dynamics and give mechanistic insight into how the global dynamics respond to perturbations in the network. The Langevin equations would take the form:

$$\begin{aligned}\frac{d}{dt}X_1 &= F_1\left(\{A_1, A_2, ..., A_N, t\}\right) \\ &\quad + \int_0^t ds K_1(t-s) X_1(s) + \eta_1(t)\end{aligned} \tag{8-23}$$

and

$$\begin{aligned}\frac{d}{dt}X_2 &= F_2\left(\{A_1, A_2, ..., A_N, t\}\right) \\ &\quad + \int_0^t ds K_2(t-s) X_2(s) + \eta_2(t)\end{aligned} \tag{8-24}$$

These equations consist of a deterministic part $F_i\left(\{A_1, A_2, ..., A_N, t\}\right)$, a noise term $\eta_i(t)$ resulting from random bath variables (faster variables that have been integrated out), and a friction (memory) term $\int_0^t ds K_i(t-s) X_i(s)$ that results from the previous interactions between the slow variables and the random bath variables.

Chapter 8.4 Conclusions

We presented an analysis that attempts to gain insight into the mechanisms that govern growth factor derived mammalian signal transduction as probed using mass spectrometry based phosphoproteomics methods. We used T=7 independent samples (taken from 7 separate stimulation conditions) of the network configuration (i.e. the relative amount of phosphorylated tyrosine at each of the N sites) to constrain the local interactions between pY sites by the measured covariance matrix. We then considered the principle of maximum entropy to derive a model of the phosphotyrosine site interaction network using only the constraints determined by the pairwise correlations that are observed in the data. A matrix of network connections was computed by taking the two-point couplings obtained from the resulting Gaussian model.

The resulting analysis reveals a surprisingly rich structure of interconnected signaling nodes. For a sufficiently high threshold value of interaction strength, a 'small world' network topology is observed in which the mean path length between two nodes in the network is much smaller than would be expected for a random network[175]. The core structure of the derived network connectivity contains many known signaling immediate and local signaling interactions (local in the sense that the activity of two of the pY sites are closely connected in the over scheme of RTK mediated signal transduction.

The model also makes several predictions regarding the network connectivity of pY sites that were hitherto not associated with the EGFR signaling network. The model therefore generates numerous biological hypotheses and it is our hope that many these

hypotheses concerning the nature of interactions in RTK signaling networks can be tested using methods such as siRNA knockdowns at particular signaling nodes. Such a model of network connectivity may then serve as a resource for the biological community in generating new hypotheses on the nature of protein-protein interaction networks in cell signaling.

Chapter 9 – Conclusions and Future Outlook

Chapter 9.1. Conclusions

It was the focus of this thesis to demonstrate, in broadly different contexts, how emergent properties from the collective interactions of the many components in molecular signaling systems. In carrying out these analyses, I was hoping to reveal some of the undoubtedly many design principles that regulate signal transduction and intercellular communication. It is my hope that this thesis will reach a broad audience; to physicists, physical scientists, and engineers who are interested in learning about how different mathematical and computational techniques, rooted in physical chemistry and statistical physics, can be applied to study biological problems; and to biologists, to open the possibility that frameworks, historically rooted in other disciplines, can sometimes provide novel biological insight and new predictive capabilities.

In the next and final section, to conclude, I will briefly discuss some future directions in signal transduction in which modeling efforts may prove useful.

Chapter 9.2. Future Outlook.

Aberrant cell growth and proliferation resulting from unregulated signaling is found in nearly half of all human cancers and numerous other diseases such as those derived from autoimmunity as in the case of diabetes[4]. Despite the enormous medical importance, even a qualitative understanding of the molecular principles that regulate signal transduction, in the networks that govern these processes, is still in its infancy. Although much of this lack of understanding is symptomatic of the inadequacy of the current technology that is available to probe these complex cellular processes, shortcomings in theoretical and computational frameworks to address these questions further compound these technological difficulties.

I will discuss three broad areas of future interest that, I believe, are both tantamount to understanding mammalian cell signaling and amenable to theoretical and computational approaches. Within each area, specific problems are posed and a preliminary discussion on how they may be addressed is considered. Connections to previous work I have carried out in my doctoral studies and have discussed in my thesis are mentioned.

Chapter 9.2.1. Development of information theoretic methods to characterize signaling dynamics derived from global phosphoproteomics data in mammalian cells

Advances in mass spectrometry have uncovered a wealth of data on the signaling dynamics in healthy and diseased cells, allowing for quantitative comparisons. For example, technologies developed in the laboratories of Mattias Mann and Forest White have enabled the real time monitoring of the trajectories of active signaling proteins measured by quantitative changes to levels of tyrosine phosphorylation on key signaling intermediates[165, 166]. Tyrosine phosphorylation is a key reversible post-translational modification utilized in mammalian cell signaling and can be used to monitor the activity state of kinases involved in signaling. In these experiments, the activities of over one hundred kinases can be measured simultaneously in real time.

The enormous number of measured degrees of freedom complicates the analysis of these data and provides the impetus to look to theory and computational modeling for representations of the data that provide predictive capabilities and intuitive physical insight.

Much of the recent modeling efforts have focused on predicting phenotypes from dimensional reduction of these data by employing linear mappings of the data onto a

phenotypic response and a subsequent principal component analysis (PCA) of the calculated correlations[181]. In such an analysis, the timecourses of each phosphorylated kinase constitute the rows of a matrix and this matrix is mapped onto a new matrix by partial least-squares regression in which the rows are taken to be different cellular conditions and columns different functional outcomes such as cell proliferation. The resulting model is then subject to a principle components analysis (PCA) and a subset of the eigenvectors obtained is used as a gauge of what elements within the network are most correlated with different responses.

Although often useful in their predictive capabilities, these techniques have limitations. First, it is not clear why a linear mapping in the data set should be presupposed to exist. Further, the eigenvectors obtained from the PCA are empirical and show only which directions are important to fitting the data. As such the method does not necessarily give any new mechanistic insight into the correlated structure within the network.

Others have approached high dimensional problems from the perspective of information theory in which the fundamental quantities of entropy and mutual information provide natural measures of correlation and order in complex systems. For example, Bialek and colleagues have derived a systematic decomposition of information content in a network by constructing correlation measures that consider entropy differences between of interacting variables and those taken independently[171, 182]. In another example, Wiggins and colleagues have derived a metric to measure the degree of network modularity by considering changes in the mutual information with respect to changes in network interactions[183]. These measures, respectively, are analogous to

irreducible correlation functions and order parameters that are used to probe structure in the framework of statistical mechanics and quantum field theory. As such, they may provide a physically insightful starting framework to think about cell signaling. It will be interesting to explore what can be understood through such information-theoretic measures and whether the computation of these measures could provide the starting point for a systematic theory that accounts for correlations in cell signaling dynamics on a strongly interacting network.

Chapter 9.2.2 The statistical physics and functional implications of multisite phosphorylation

The utilization of multiple phosphorylation sites in regulating a biological response is ubiquitous in signal transduction, especially in the regulation of growth in mammalian cells[184]. The functional consequences of a protein having multiple phosphorylation sites, as opposed to activation resulting from a single phosphorylation event, are many[1, 109, 185]. However, a complete understanding of how multiple reversible phosphorylations are used synergistically to regulate a biological response is not understood.

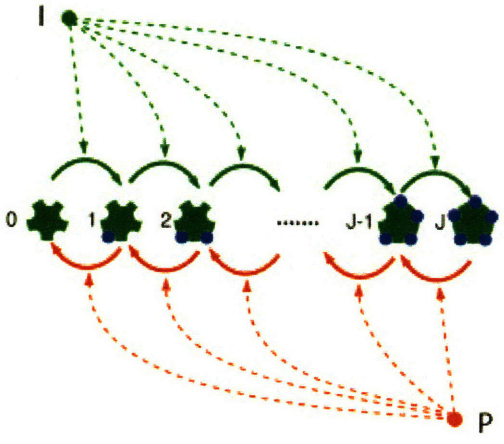


Figure 9-1. **Schematic of Multisite Phosphorylation.**

The activity of a protein is often determined by its phosphorylation state on multiple sites. In this schematic, a single protein can be phosphorylated by I and dephosphorylated by P at a total of J different sites. Figure adapted from [1].

If each site contributes a redundant binding site as is often the case, then one consequence of multiple phosphorylation sites is to change the probability that a protein, once dissociated, immediately rebinds to its target[109, 185]. The consequences of such effect has been studied theoretically[185], and it is predicted that the kinetics of dissociation of a protein bound to a protein with multiple phosphorylation sites can exhibit marked deviations from exponential kinetics. In this regime, the dissociation kinetics have a large tail and proteins remain bound for times much longer than its characteristic time for unbinding τ_{off}. How such effects may influence the dynamics of downstream signaling as well as the association of other proteins on the same multiply phosphorylated protein has not been explored. It would be interesting to develop an analytical theory that can account for multisite phosphorylation, substrate rebinding, and the simultaneous interaction of other components in this signaling pathway.

Another intriguing possibility, for how multisite phosphorylation might be utilized, arises upon noting that Src-homology 2 recognition domains[186], that recognize phosphorylated tyrosine, often are present on proteins in duplicates and can potentially bind to their target in tandem. This two-site interaction allows for the possibility of cross-linking multiple proteins through their multiple phosphorylation sites, ultimately resulting in the formation of large aggregates of signaling proteins. It is not well understood what consequences in signaling arise from the effects of cross linking and cluster formation. For example, excluded volume interactions may play an important role in defining the structure of these clusters and potentially allow for long-range density correlations in the locations of phosphorylation sites. Such effects may be analogous to

the physics of excluded volume interactions that give rise to long range density correlations in well-studied models of diffusion limited aggregation [187]. Such correlations could have dramatics effects on signal propagation.

Finally, another potential way to study the physics of multisite phosphorylation is to consider a mapping between a sequence of N phosphorylation sites and the statistical mechanics of a class of one dimensional Ising models. If the sequence of phosphorylation sites is phosphorylated and dephosphorylated (represented by up and down spins) in sequential order, then the statistics of multisite phosphorylation maps exactly onto a one dimensional kinetically constrained Ising model known as the East model[188] in which the ratio of phosphorylation and desphosphorylation rates acts as an effective temperature.

Once such a mapping is made, several analogies can be explored. For example, it has been recently shown that these models at certain temperatures display rich, correlated, dynamical behavior in the trajectory space of the model[189] – although space is unstructured at any time point, long range correlations persist when looking different nearby time points. A detailed consideration of this potential mapping has not been explored and may yield interesting connections to how phosphorylation patterns on single proteins are regulated. Information, in the form of spatio-temporal correlated regions of phosphorylation space, may yield new biological insights.

Chapter 9.2.3 Modeling the integration of nutrient, stress, and growth-factor derived signals by the PI3K/AKT pathway and mTORC1/2 signaling complexes

The PI3K/AKT pathway integrates signals derived from growth-factors, nutrients, and cellular stresses to allow for a multitude of cellular responses including changes in

metabolism and the upregulation of genes necessary for growth and proliferation[6]. A key player in this pathway is the mammalian target of rapamycin (mTOR) protein. Recent biochemical evidence[190] suggests that the pathway integrates signals by way of two distinct signaling complexes, mTORC1 and mTORC2; both complexes share the same (mTOR) protein. mTOR interacts both upstream and downstream of AKT; AKT is a kinase that functions as a master regulator having over 100 substrates. A schematic of the current model of AKT regulation is shown in Fig. 9-2.

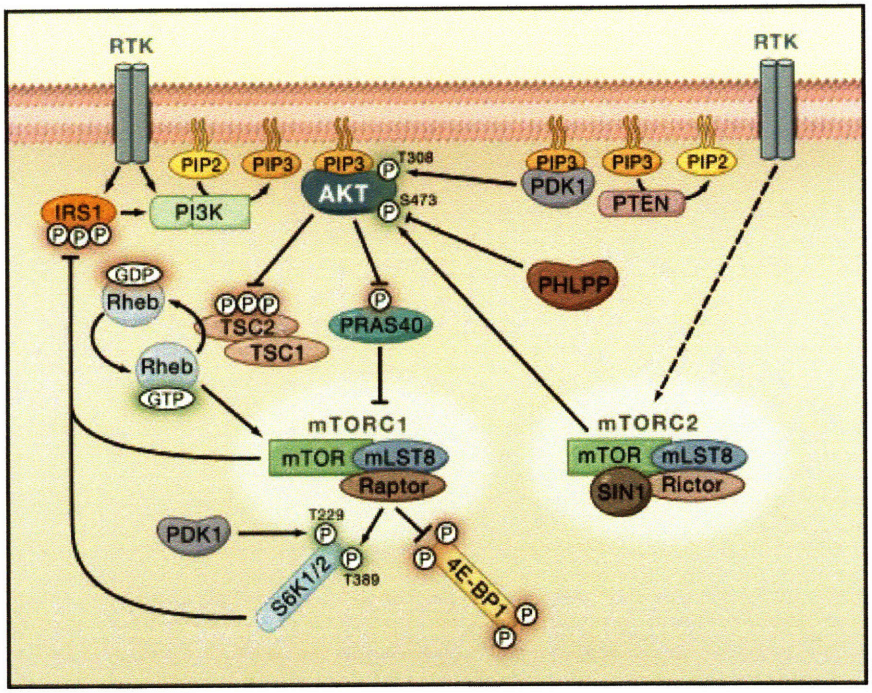


Figure 9-2. **Current model of how PI3K/AKT signaling is regulated.**

mTORC1/2 complexes interact both upstream and downstream of AKT. AKT controls many cellular processes such as cell growth, proliferation, apoptosis, migration, and metabolism. Schematic is taken from [6].

From the diagram, one immediate observation is apparent: mTOR and its complexes provide multiple positive and negative interactions resulting in

phosphorylations in AKT. While the mTORC2 complex directly phosphorylates the ser473 residue in AKT, the mTORC1 complex provides multiple indirect feedback interactions via the activity of the adapter protein IRS1. A further complication arises from the scaffolding nature of mTOR in regulating the mTORC1 and mTORC2 complexes. The assembly of multiprotein complexes by scaffolding proteins has been to shown to have diverse consequences on signal propagation[12, 13]. Thus, the expression level of mTOR becomes essential in determining the activity of both complexes. Furthermore, a recent theoretical study predicts that the concentration of scaffolding proteins influences the dynamics of kinase activation diverse ways[13]. Depending on the concentration, it is predicted that the activity of the mTORC1 and mTORC2 complexes can be either be temporally distributed with a single exponential time scale or broadly distributed across many time scales. How these predicted different waiting time distributions of kinase activity influence the many positive and negative interactions present both upstream and downstream of AKT is not understood. A theoretical model that couples different temporal profiles of mTORC1 and mTORC2 activity into the regulation of AKT may provide some insight into this very complicated question.

Another obstacle that is currently preventing a detailed mechanistic understanding of signal integration by this pathway is the lack of biochemical data that details the many interactions in this pathway. Gross uncertainties in both parameter space and connectivity of the network preclude any single biochemical model of the pathway. Instead, a statistical ensemble approach in which model parameters and the network connectivity fluctuate may be a more appropriate treatment[191, 192]. Knowledge of the distribution of biochemical models that accurately captures the known or possible

phenomenology contained within the PI3K/AKT pathway would likely yield fruitful dividends especially in the case of parsing the effects of combinations of the many positive and negative regulatory interactions conferred by the mTORC1/2 complexes.

References

[1] S. Krishnamurthy *et al.*, Biology Direct **2** (2007).
[2] J. W. Locasale, and A. K. Chakraborty, In review (PLoS Computatoinal Biology (2008).
[3] J. W. Locasale, and A. K. Chakraborty, in preparation (2008).
[4] R. J. Shaw, and L. C. Cantley, Nature **441**, 424 (2006).
[5] L. C. Cantley, Science **296**, 1655 (2002).
[6] B. D. Manning, and L. C. Cantley, Cell **129**, 1261 (2007).
[7] T. Pawson, Cell **116**, 191 (2004).
[8] T. Pawson, Current Opinion in Cell Biology **19**, 112 (2007).
[9] T. Pawson, and J. D. Scott, Trends in Biochemical Sciences **30**, 286 (2005).
[10] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman & Hall, 2007).
[11] T. M. Cover, and J. A. Thomas, *Elements of Information Theory* (Wiley, 2006).
[12] J. W. Locasale, A. S. Shaw, and A. K. Chakraborty, Proceedings of the National Academy of Sciences of the United States of America **104**, 13307 (2007).
[13] J. W. Locasale, and A. K. Chakraborty, Plos Computational Biology **(in press)** (2008).
[14] J. W. Locasale, arXiv:0802.2683 [qbio.MN] (2008) (2008).
[15] J. W. Locasale, Journal of Chemical Physics **128**, 10 (2008).
[16] J. W. Locasale, PLoS ONE **2**, e627 (2007).
[17] S. Cemerski *et al.*, Immunity **26**, 345 (2007).
[18] B. e. a. Alberts, *Molecular biology of the cell* (Garland Science, New York, NY, 2002).
[19] L. F. Chang, and M. Karin, Nature **410**, 37 (2001).
[20] M. S. Qi, and E. A. Elion, Journal of Cell Science **118**, 3569 (2005).
[21] J. E. Ferrell, and E. M. Machleder, Science **280**, 895 (1998).
[22] P. S. Swain, and E. D. Siggia, Biophysical Journal **82**, 2928 (2002).
[23] B. N. Kholodenko, European Journal of Biochemistry **267**, 1583 (2000).
[24] W. R. Burack, and A. S. Shaw, Current Opinion in Cell Biology **12**, 211 (2000).
[25] W. Kolch, Nature Reviews Molecular Cell Biology **6**, 827 (2005).
[26] D. K. Morrison, and R. J. Davis, Annual Review of Cell and Developmental Biology **19**, 91 (2003).
[27] N. Dard, and M. Peter, Bioessays **28**, 146 (2006).
[28] D. Bray, Annual Review of Biophysics and Biomolecular Structure **27**, 59 (1998).
[29] A. Levchenko, J. Bruck, and P. W. Sternberg, Proceedings of the National Academy of Sciences of the United States of America **97**, 5818 (2000).
[30] R. P. Bhattacharyya *et al.*, Science **311**, 822 (2006).
[31] R. Heinrich, B. G. Neel, and T. A. Rapoport, Molecular Cell **9**, 957 (2002).
[32] J. E. Ferrell, Trends in Biochemical Sciences **21**, 460 (1996).
[33] M. Arrio-Dupont *et al.*, Biophysical Journal **78**, 901 (2000).
[34] A. Goldbeter, and D. E. Koshland, Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **78**, 6840 (1981).
[35] D. Frenkel, and B. Smit, *Understanding Molecular Simulation. From Algorithms to Applications* (Academic Press: Boston, MA, 2002).

[36] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, NY, 1987).
[37] A. S. Shaw.
[38] A. von Kriegsheim *et al.*, Nature Cell Biology **8**, 1011 (2006).
[39] M. K. Dougherty *et al.*, Molecular Cell **17**, 215 (2005).
[40] D. A. Ritt, I. O. Daar, and D. K. Morrison, in *Regulators and Effectors of Small Gtpases: Ras Family*2006), pp. 224.
[41] A. Nguyen *et al.*, Molecular and Cellular Biology **22**, 3035 (2002).
[42] K. Scott, and C. S. Zuker, Nature **395**, 805 (1998).
[43] M. Bentires-Alj *et al.*, Nature Medicine **12**, 114 (2006).
[44] R. L. Kortum *et al.*, Molecular and Cellular Biology **26**, 2202 (2006).
[45] S. H. Park, A. Zarrinpar, and W. A. Lim, Science **299**, 1061 (2003).
[46] A. Harding *et al.*, Current Biology **15**, 869 (2005).
[47] S. Ghaemmaghami *et al.*, Nature **425**, 737 (2003).
[48] W. R. Burack, and T. W. Sturgill, Biochemistry **36**, 5929 (1997).
[49] L. Leibler, Macromolecules **13**, 1602 (1980).
[50] F. H. Stillinger, and A. Bennaim, Journal of Chemical Physics **74**, 2510 (1981).
[51] S. J. McQuarrie DA, *Physical Chemistry: A Molecular Approach* (University Science Books, 1997).
[52] T. Pawson, and J. D. Scott, Science **278**, 2075 (1997).
[53] K. L. Dodge-Kafka *et al.*, Nature **437**, 574 (2005).
[54] R. L. Kortum, and R. E. Lewis, Molecular and Cellular Biology **24**, 4407 (2004).
[55] R. L. Kortum *et al.*, Molecular and Cellular Biology **25**, 7592 (2005).
[56] F. D. Smith, L. K. Langeberg, and J. D. Scott, Trends in Biochemical Sciences **31**, 316 (2006).
[57] C. J. Bashor *et al.*, Science **319**, 1539 (2008).
[58] C. J. Marshall, Cell **80**, 179 (1995).
[59] L. O. Murphy, and J. Blenis, Trends in Biochemical Sciences **31**, 268 (2006).
[60] L. O. Murphy *et al.*, Nature Cell Biology **4**, 556 (2002).
[61] R. E. Dolmetsch *et al.*, Nature **386**, 855 (1997).
[62] L. F. Chen *et al.*, Science **293**, 1653 (2001).
[63] S. Paul *et al.*, Nature Neuroscience **6**, 34 (2003).
[64] Y. R. Chen *et al.*, Journal of Biological Chemistry **271**, 31929 (1996).
[65] M. A. Daniels *et al.*, Nature **444**, 724 (2006).
[66] S. Sasagawa *et al.*, Nature Cell Biology **7**, 365 (2005).
[67] B. Schoeberl *et al.*, Nature Biotechnology **20**, 370 (2002).
[68] S. C. Kao *et al.*, Journal of Biological Chemistry **276**, 18169 (2001).
[69] S. Boykevisch *et al.*, Current Biology **16**, 2173 (2006).
[70] S. D. M. Santos, P. J. Verveer, and P. I. H. Bastiaens, Nature Cell Biology **9**, 324 (2007).
[71] A. Kiyatkin *et al.*, Journal of Biological Chemistry **281**, 19925 (2006).
[72] B. N. Kholodenko, Nature Reviews Molecular Cell Biology **7**, 165 (2006).
[73] T. Lu *et al.*, Proceedings of the National Academy of Sciences of the United States of America **103**, 16752 (2006).
[74] Y. H. Lan, and G. A. Papoian, Physical Review Letters **98** (2007).

[75] S. Redner, *A Guide To First Passage Processes* (Cambridge University Press, 2001).
[76] D. A. Lauffenburger, and J. A. Linderman, *Receptors: Models for Binding, Trafficking, and Signaling* (Oxford University Press, 1993).
[77] S. Tanase-Nicola, P. B. Warren, and P. R. ten Wolde, Physical Review Letters **97** (2006).
[78] J. S. van Zon *et al.*, Biophysical Journal **91**, 4350 (2006).
[79] P. B. Warren, S. Tanase-Nicola, and P. R. ten Wolde, Journal of Chemical Physics **125** (2006).
[80] H. P. Lu, L. Y. Xun, and X. S. Xie, Science **282**, 1877 (1998).
[81] J. Yu *et al.*, Science **311**, 1600 (2006).
[82] B. D. Slaughter, J. W. Schwartz, and R. Li, Proceedings of the National Academy of Sciences of the United States of America **104**, 20320 (2007).
[83] D. S. Frenkel, B, *Understanding Molecular Simulation. From Algorithms to Applications* (Academic Press: Boston, MA, 2002).
[84] A. R. Asthagiri, and D. A. Lauffenburger, Biotechnology Progress **17**, 227 (2001).
[85] M. Thattai, and A. van Oudenaarden, Biophysical Journal **82**, 2943 (2002).
[86] M. Chaves, E. A. Sontag, and R. J. Dinerstein, Journal of Physical Chemistry B **108**, 15311 (2004).
[87] K. C. El Kasmi *et al.*, Journal of Immunology **177**, 7880 (2006).
[88] M. Behar *et al.*, Biophysical Journal **93**, 806 (2007).
[89] M. Behar, H. G. Dohlman, and T. C. Elston, Proceedings of the National Academy of Sciences of the United States of America **104**, 16146 (2007).
[90] L. Bardwell *et al.*, Biophysical Journal **92**, 3425 (2007).
[91] J. J. Hornberg *et al.*, Oncogene **24**, 5533 (2005).
[92] I. Amit *et al.*, Nature Genetics **39**, 503 (2007).
[93] A. Hoffmann *et al.*, Science **298**, 1241 (2002).
[94] S. L. Werner, D. Barken, and A. Hoffmann, Science **309**, 1857 (2005).
[95] M. Ptashne, *Genes and Signals* (Cold Spring Harbor Laboratory Press, 2001).
[96] P. Cohen, Trends in Biochemical Sciences **25**, 596 (2000).
[97] X. J. Yang, Oncogene **24**, 1653 (2005).
[98] M. A. Pufall *et al.*, Science **309**, 142 (2005).
[99] P. Miller *et al.*, Plos Biology **3**, 705 (2005).
[100] K. Haglund, and I. Dikic, Embo Journal **24**, 3353 (2005).
[101] J. Gunawardena, Proceedings of the National Academy of Sciences of the United States of America **102**, 14617 (2005).
[102] F. Stegmeier *et al.*, Nature **446**, 876 (2007).
[103] H. C. Vodermaier, Current Biology **14**, R787 (2004).
[104] N. E. Buchler, U. Gerland, and T. Hwa, Proceedings of the National Academy of Sciences of the United States of America **102**, 9559 (2005).
[105] K. Nasmyth, Cell **120**, 739 (2005).
[106] M. Rape, and M. W. Kirschner, Nature **432**, 588 (2004).
[107] S. K. Reddy *et al.*, Nature **446**, 921 (2007).
[108] M. Rape, S. K. Reddy, and M. W. Kirschner, Cell **124**, 89 (2006).
[109] P. Nash *et al.*, Nature **414**, 514 (2001).

[110] H. J. Dyson, and P. E. Wright, Nature Reviews Molecular Cell Biology **6**, 197 (2005).
[111] S. Orlicky *et al.*, Cell **112**, 243 (2003).
[112] P. Klein, T. Pawson, and M. Tyers, Current Biology **13**, 1669 (2003).
[113] A. Levchenko, Current Biology **13**, R876 (2003).
[114] H. C. Berg, and E. M. Purcell, Biophysical Journal **20**, 193 (1977).
[115] B. C. Lagerholm, and N. L. Thompson, Biophysical Journal **74**, 1215 (1998).
[116] M. Gopalakrishnan *et al.*, Biophysical Journal **89**, 3686 (2005).
[117] M. Gopalakrishnan *et al.*, European Biophysics Journal with Biophysics Letters **34**, 943 (2005).
[118] B. Goldstein, and M. Dembo, Biophysical Journal **68**, 1222 (1995).
[119] D. Shoup, and A. Szabo, Biophysical Journal **40**, 33 (1982).
[120] S. Y. Shvartsman *et al.*, Biophysical Journal **81**, 1854 (2001).
[121] S. S. Andrews, Physical Biology **2**, 111 (2005).
[122] B. C. Lagerholm, and N. L. Thompson, Journal of Physical Chemistry B **104**, 863 (2000).
[123] A. M. Lieto, B. C. Lagerholm, and N. L. Thompson, Langmuir **19**, 1782 (2003).
[124] M. D. Levin, T. S. Shimizu, and D. Bray, Biophysical Journal **82**, 1809 (2002).
[125] R. Zwanzig, Proceedings of the National Academy of Sciences of the United States of America **87**, 5856 (1990).
[126] C. Salazar, and T. Hofer, Febs Journal **274**, 1046 (2007).
[127] M. M. Borg, T. Pawso, T. Tyers, M. Forman-Kay, JD. Chan, HS., Proc Natl Acad Sci **104**, 9650 (2007).
[128] O. Flomenbom, and R. J. Silbey, Proceedings of the National Academy of Sciences of the United States of America **103**, 13897 (2006).
[129] S. C. Kou *et al.*, Journal of Physical Chemistry B **109**, 19068 (2005).
[130] B. Kusse, and E. Westwig, *Mathematical Physics: Applied Mathematics for Scientists and Engineers* (John Wiley & Sons, Inc., 1998).
[131] D. J. Tschumperlin *et al.*, Nature **429**, 83 (2004).
[132] K. A. Janes *et al.*, Cell **124**, 1225 (2006).
[133] M. Hermanson *et al.*, Cancer Research **52**, 3213 (1992).
[134] P. Lenz, and P. S. Swain, Current Biology **16**, 2150 (2006).
[135] S. Kim, and J. Ferrell, Cell **128**, 1133 (2007).
[136] R. G. Posner, C. Wofsy, and B. Goldstein, Mathematical Biosciences **126**, 171 (1995).
[137] M. Doi, and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, 1988).
[138] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier Science Publishing, 2001).
[139] A. Abramowitz, and I. Stegun, *Handbook of Mathematical Functions* (Dover Publications, 1965).
[140] C. Sumen *et al.*, Immunity **21**, 315 (2004).
[141] T. R. Mempel, S. E. Henrickson, and U. H. von Andrian, Nature **427**, 154 (2004).
[142] M. Faroudi *et al.*, Journal of Immunology **171**, 1128 (2003).
[143] I. Munitic, P. E. Ryan, and J. D. Ashwell, Journal of Immunology **174**, 4010 (2005).

[144] K. H. Lee *et al.*, Science **302**, 1218 (2003).
[145] G. Altan-Bonnet, and R. N. Germain, Plos Biology **3**, 1925 (2005).
[146] Q. J. Li *et al.*, Nature Immunology **5**, 791 (2004).
[147] D. C. Wylie, J. Das, and A. K. Chakraborty, Proceedings of the National Academy of Sciences of the United States of America **104**, 5533 (2007).
[148] C. Dong, R. J. Davis, and R. A. Flavell, Annual Review of Immunology **20**, 55 (2002).
[149] M. A. Goldsmith, and A. Weiss, Science **240**, 1029 (1988).
[150] A. E. Schade, and A. D. Levine, Journal of Immunology **172**, 5828 (2004).
[151] C. Rosette *et al.*, Immunity **15**, 59 (2001).
[152] C. W. Arendt *et al.*, Current Opinion in Immunology **14**, 323 (2002).
[153] M. M. Winslow, J. R. Neilson, and G. R. Crabtree, Current Opinion in Immunology **15**, 299 (2003).
[154] J. E. Ferrell, Current Opinion in Cell Biology **14**, 140 (2002).
[155] C. Janeway, *Immunobiology* (Garland Science, 2004).
[156] J. P. MacKeigan *et al.*, Molecular and Cellular Biology **25**, 4676 (2005).
[157] M. Gao *et al.*, Science **306**, 271 (2004).
[158] M. Reth, and T. Brummer, Nature Reviews Immunology **4**, 269 (2004).
[159] W. Xiong, and J. E. Ferrell, Nature **426**, 460 (2003).
[160] T. Yokosuka *et al.*, Nature Immunology **6**, 1253 (2005).
[161] A. Grakoui *et al.*, Science **285**, 221 (1999).
[162] D. T. Gillespie, Journal of Computational Physics **22**, 403 (1976).
[163] M. A. Markiewicz *et al.*, J Immunol **175**, 2825 (2005).
[164] M. Krogsgaard *et al.*, Nature **434**, 238 (2005).
[165] Y. Zhang *et al.*, Molecular & Cellular Proteomics **4**, 1240 (2005).
[166] J. V. Olsen *et al.*, Cell **127**, 635 (2006).
[167] A. Wolf-Yadlin *et al.*, Molecular Systems Biology (2006).
[168] A. Wolf-Yadlin *et al.*, Proceedings of the National Academy of Sciences of the United States of America **104**, 5860 (2007).
[169] K. A. Janes, and M. B. Yaffe, Nature Reviews Molecular Cell Biology **7**, 820 (2006).
[170] T. R. Lezon *et al.*, Proceedings of the National Academy of Sciences of the United States of America **103**, 19033 (2006).
[171] E. Schneidman *et al.*, Nature **440**, 1007 (2006).
[172] A. J. Chorin, A. P. Kast, and R. Kupferman, Proceedings of the National Academy of Sciences of the United States of America **95**, 4094 (1998).
[173] A. J. Chorin, O. H. Hald, and R. Kupferman, Proceedings of the National Academy of Sciences of the United States of America **97**, 2968 (2000).
[174] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, 2001), p. 240 pages.
[175] R. Albert, and A. L. Barabasi, Reviews of Modern Physics **74**, 47 (2002).
[176] S. Audebert *et al.*, Current Biology **14**, 987 (2004).
[177] Y. Shimomura *et al.*, Journal of Dermatological Science **31**, 43 (2003).
[178] R. Kittler *et al.*, Nature **432**, 1036 (2004).
[179] D. A. Six, and E. A. Dennis, Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids **1488**, 1 (2000).

[180] J. P. Pin, T. Galvez, and L. Prezeau, Pharmacology & Therapeutics **98**, 325 (2003).
[181] N. Kumar *et al.*, Plos Computational Biology **3**, 35 (2007).
[182] E. Schneidman *et al.*, Physical Review Letters **91** (2003).
[183] E. Ziv, M. Middendorf, and C. H. Wiggins, Physical Review E **71** (2005).
[184] P. J. Roach, Journal of Biological Chemistry **266**, 14139 (1991).
[185] J. W. Locasale, Journal of Chemical Physics (in press) (2008).
[186] B. T. Seet *et al.*, Nature Reviews Molecular Cell Biology **7**, 473 (2006).
[187] T. A. Witten, and L. M. Sander, Physical Review Letters **47**, 1400 (1981).
[188] F. Ritort, and P. Sollich, Advances in Physics **52**, 219 (2003).
[189] J. P. Garrahan, and D. Chandler, Physical Review Letters **89** (2002).
[190] D. D. Sarbassov *et al.*, Science **307**, 1098 (2005).
[191] K. S. Brown, and J. P. Sethna, Physical Review E **68** (2003).
[192] L. Kuepfer *et al.*, Nature Biotechnology **25**, 1001 (2007).